WIDER Working Paper 2021/183

# The gender productivity gap

## Evidence from the informal sector in India

Ira N. Gang,[1] Rajesh Raj Natarajan,[2] Kunal Sen,[3] and Myeong-Su Yun[4,*]

December 2021

**Abstract:** We examine the patterns and correlates of the productivity gap between male-owned and female-owned firms for informal enterprises in India. Female-owned firms are on average 45 per cent less productive than male-owned firms, with the clearest productivity gaps observed at the lower end of the productivity distribution. Using decomposition methods, we find that about 73 per cent of the productivity gap can be explained by structural effect, with the remainder being due to differences in observable characteristics as captured by composition effect. We also find that among observable characteristics, the most important contributing factors explaining the gender productivity gap are firm characteristics, such as firm size, age of the firm, assistance from the government, registration with state authorities, working on a contract basis, and maintaining accounts. Male-owned firms are more advantaged in these characteristics than female-owned firms.

**Key words:** gender, productivity gap, India, decomposition methods, informal sector

**JEL classification:** J17, O17, O53

**Acknowledgements:** This paper uses unit-level data on informal firms from the National Sample Survey Office, and access to the same is gratefully acknowledged. The usual disclaimers apply.

---

[1] Rutgers University, New Jersey, USA; RWI, Essen; CReAM, London; IOS, Regensburg; GLO, Essen; and IZA, Bonn; [2] Department of Economics, Sikkim University, Sikkim, India; [3] UNU-WIDER, Helsinki, Finland; [4] Department of Economics, Inha University, South Korea; GLO, Essen; and IZA, Bonn; * corresponding author: Myeong-Su Yun, msyun@inha.ac.kr.

This study is published within the UNU-WIDER project Women's work – routes to economic and social empowerment.



Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-123-5

https://doi.org/10.35188/UNU-WIDER/2021/123-5

Typescript prepared by Ayesha Chari.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

## 1 Introduction

A long-standing literature has pointed out deep and widespread gender inequalities in labour market outcomes (see World Bank 2012). However, the focus of the literature on gender inequality in labour markets has been on gender wage gaps and gender differences in labour force participation (see Bertrand 2020; Borrowman and Klasen 2020). There has been less interest in the literature on gender differences in firm productivity, especially for small and micro enterprises. Yet, as has been documented in the previous literature, female micro entrepreneurs face significant disadvantages compared with male micro entrepreneurs (Batista et al. 2021).

This paper examines the correlates of gender differences in the productivity of male-owned and female-owned firms in the informal sector in India. In particular, we analyse gender-based differences in firm productivity and the factors contributing to it. Our main objective is to assess whether a performance gap between male-owned and female-owned enterprises, as observed in the majority of the existing studies, is present in informal sector enterprises too. Additionally, the paper also attempts to understand the factors contributing to this gap.

Focusing on the informal sector is important, as much of economic activity is in this heterogeneous sector. Informal firms are typically household units, survivalist in nature, with limited growth prospects (Grimm et al. 2012). It is a sector characterized by both entrepreneurial activity and serves as an employer of last resort. India's informal sector includes manufacturing firms and unincorporated non-agricultural enterprises, making up approximately 75 per cent of manufacturing employment and 17 per cent of manufacturing output (Raj and Sen 2016a). Our strategy lies in taking advantage of nationwide data on unorganized (informal) enterprises in India collected in 2015–16 by India's National Sample Survey Organization (NSSO).

If female-run enterprises disproportionately exhibit lower productivity, then it is likely the women running and working in these firms will be less successful in escaping poverty and in improving their living standards than the men in the sector. In decomposing gender difference in firm productivity in India's informal sector, we find that the gender differences are substantially large among the less productive firms and relatively small among highly productive firms. How to address this disparity in outcomes, this source of social tensions, is a question faced by society and policymakers alike.

We find that on average, male-owned firms are 45 per cent more productive than female-owned firms. However, this productivity gap by gender is particularly noticeable at the bottom and middle parts of the productivity distribution. This implies that there is relatively no gender productivity gap for the most productive firms in the informal sector in India. On the other hand, the smaller of the female-owned firms are at a productivity disadvantage compared with their male-owned counterparts. Given that most of the owners of the smaller informal enterprises are part of the working poor in India, this has clear negative implications for gender equality in living standards for small and micro entrepreneurs in India's informal sector.

We proceed by examining the correlates of the gender productivity gap using Oaxaca and recentred influence function (RIF) decomposition methods. We find that about 73 per cent of the productivity gap can be explained by structural effect and the remaining gap explained by differences in observable characteristics (composition effect). We also find that among observable characteristics, the most important contributing factors explaining the gender productivity gap are firm characteristics for both composition and structural effects. Male-owned firms are more advantaged in firm characteristics such as firm size, age of the firm, assistance from the government, registration with state authorities, working on a contract basis, and maintaining accounts as well as earning more for the same characteristics. Regional and sectoral effects also matter in explaining structural effect.

This paper is divided into six sections. In Section 2, we provide review of the literature on the determinants of firm productivity, with specific reference to India. Section 3 briefly describes our methodology. In Section 4, we carefully present our data, including our sample selection, variable construction, descriptive statistics, and figures capturing the productivity differences between male-owned and female-owned firms. This firmly sets the stage for our empirical analysis and discussion in Section 5, in particular our decision to examine the sources of gender productivity gaps not only at the mean but also across the distribution using both Oaxaca and RIF decompositions. Section 6 concludes.

## 2 Related literature

In this section, we provide a brief review of the literature on the determinants of firm productivity, with specific reference to India. Three broad sets of factors have been highlighted: (i) firm characteristics, (ii) firm constraints, and (iii) social groups relevant in the Indian context, in particular, caste.

### 2.1 Firm characteristics

In the resource-based view of firm capabilities, where the firm is not just an administrative unit but a collection of productive resources (Penrose 1959), the capabilities of the managers of firms (who, in the case of informal firms, are also the owners) are key to attaining competitive advantage and to firm expansion. Studies have used different indicators to represent firm capabilities. Indicators such as investment in research and development or purchase of specialized machinery are less applicable as measures of firm capability in the informal sector (Sher and Yang 2005; Yang and Huang 2005). A key source of information on technology and marketing for owners of informal firms is other firms, often in the formal sector, with whom they have sub-contracting arrangements (Berry et al. 2002). In our empirical analysis of the gender differences in firm productivity, we took linkage (sub-contracting) as an explanatory variable.

Another indicator of firm capabilities in the informal sector is the maintenance of accounts by a small informal firm, which may allow the owner/manager of the firm to access external finance via the presentation of these accounts to bank managers and help overcome the constraints to their expansion. Sound accounting practices are considered an important factor driving firm growth (Acar 1993). The registration of firms under a given act or authority of the government also provides a proxy for firm capabilities as registered firms are able to access specialized training and acquire knowledge compared with non-registered firms. Sharma (2014) finds that registration leads to 32 per cent gain in sales per employee and 56 per cent gain in value added per employee for firms in the small-scale sector. There exists substantial evidence on the positive role of locational factors on firm productivity. Urban firms may be more productive than rural firms because of agglomeration benefits and access to larger markets (Bigsten and Gebreeyesus 2007; Liedholm and Mead 1999).

Considering firm size, studies have shown that the larger the firm is, the better performance the firm enjoys (Ayyagari et al. 2014; Poschke 2018). Economies of scale is one of the critical factors, according to these studies, explaining the dominant performance of large firms over small firms. Some scholars have found opposite patterns for the relationship between firm size and growth

(Bollard et al. 2014; Goedhuys and Sleuwaegen 2000; Sleuwaegen and Goedhuys 2002). Existing literature also shows a correlation between firm age and performance. In this case too, studies have produced conflicting results. Those finding a negative relationship between firm age and growth suggest decreasing returns to learning over time as one of the major reasons (Calvo 2006; Ericson and Pakes 1995; Fariñas and Moreno 2000). On the other hand, studies attribute the positive relationship between firm age and performance to older firms being able to obtain more resources, such as information, experience, networks, access to finance, and better reputation (Autio 2005; Coad et al. 2013). The cost of capital tends to decline as firms get older (Hadlock and Pierce 2010) and the older firms encounter lower plant failure rates (Dunne et al. 1989).

Several governments, especially those in developing countries, have devised many programmes to provide technical assistance to small firms. Such assistance to firms can act as a catalyst for firm growth and productivity. Raj and Sasidharan (2014) show that firms receiving assistance are likely to encounter faster firm growth.

## 2.2 Firm constraints

The next set of factors we consider are those that act as constraints to increasing firm productivity. Access to and cost of finance are among the factors that determine the ability of a firm to become more productive (Binks and Ennew 1996; Oliveira and Fortunato 2006; Rajan and Zingales 1998). Firms that face financial constraints are less likely to invest in fixed assets (Ojah et al. 2010; Winker 1999) and, hence, will be less productive. Better provision of infrastructure is another factor that is likely to influence firm performance. Lack of access to infrastructure, such as access to electricity, is one of the serious constraints faced by the firms in the informal sector, which has a direct bearing on the growth and productivity performance of firms in the sector (Raj and Sen 2016a, 2016b).

## 2.3 Social group

The importance of social group in the business economy is highlighted by many studies (Mosse 2018). In terms of ownership of businesses, there exists a wide disparity among the upper castes and the marginalized communities (Thorat and Sadana 2009). This is clearly evident from the decline in the proportion of units owned by marginalized groups such as Scheduled Castes (SCs) and Scheduled Tribes (STs) (Deshpande and Sharma 2013). Studies have shown that the caste of the entrepreneurs determines their entry into several sectors (Harriss-White et al. 2014). Due to their limited access to capital or collateral, infrastructure, raw materials, and markets controlled by other castes, marginalized groups ended up running survival-oriented businesses rather than entrepreneurial businesses (Deshpande and Sharma 2016; Guérin et al. 2015; Harriss-White et al. 2014). Such caste discrimination has also influenced the performance of firms owned by entrepreneurs from marginalized groups. A study by Goraya (2019) shows that the misallocation of capital across castes leads to 43 and 34 per cent of total factor productivity and gross domestic product losses, respectively. In a study using the nationally representative data for 2004–05, Deshpande and Sharma (2016) show significant caste gap in earnings from self-employment at the lower end of the distribution. Existence of such social barriers have also influenced the performance of micro and small firms in the informal sector. For instance, using the NSSO dataset for informal sector enterprises, Raj and Sen (2016a, 2016b) show that enterprises owned by members of disadvantaged social groups such as the SCs and STs are less productive than those headed by the Other Backward Classes (OBCs) and the General category.

## 3 Methodology

The main objective of this paper is to understand the factors contributing to the gap in productivity between firms owned and operated by male and female entrepreneurs. We use Oaxaca decomposition approaches to probe the sources of this productivity differential (see Jann 2008; Rios-Avila 2020). As a first step, we perform the decomposition at the mean. Specifically, we begin by estimating a model of determinants of (log) labour productivity for male-owned firms and female-owned firms as follows:

$$lnP_{L_{eg}} = Z'_{eg}\alpha_g + \omega_{eg}; e = 1, \dots\dots, n; g = m, f \quad (1)$$

where $e$ denotes the enterprise; $g$ the gender of the firm owner, male ($m$) or female ($f$); $lnP_{L_{eg}}$ is the logarithm of labour productivity; $Z'_{eg}$ is the vector of firm-level attributes that are likely to affect labour productivity; and $\omega_{eg}$ is the random error term. Under the assumption of linearity, zero mean and constant variance, Equation (1) is estimated using ordinary least square.

We then arrive at the difference in the expected labour productivity values of male-owned and female-owned firms ($D_p$) from the estimates of Equation (1) separately for male and female managers. Symbolically,

$$D_p = \overline{lnP_{L_m}} - \overline{lnP_{L_f}} = (\bar{Z}'_m\hat{\alpha}_m) - (\bar{Z}'_f\hat{\alpha}_f) \quad (2)$$

As stated in Oaxaca (2007), it is important to decompose the gender gap in productivity into the component of the gap attributable to the differences in observable characteristics (composition effect) and to differences in returns to coefficients (structural effect). We implement this as follows:

$$D_p = (\bar{Z}'_m\hat{\alpha}_m) - (\bar{Z}'_f\hat{\alpha}_f) = (\bar{Z}_m - \bar{Z}_f)'\hat{\alpha}_m + \bar{Z}'_f(\hat{\alpha}_m - \hat{\alpha}_f) \quad (3)$$

where $\hat{\alpha}_g$ is the estimated value of $\alpha_g$. The first term on the right-hand side of Equation (3), namely, $(\bar{Z}_m - \bar{Z}_f)'\hat{\alpha}_m$, is the explained part of the gender gap in productivity, the part that captures the gap explained by the differences in observed characteristics at the mean, weighted by the coefficients attributable to male-owned firms ($\hat{\alpha}_m$).[1] In this paper, we call this composition effect. The second term, namely, $\bar{Z}'_f(\hat{\alpha}_m - \hat{\alpha}_f)$, is the unexplained part, which captures the productivity differences that are not explained by the observed predictors. This component might represent the heterogeneous response of covariates by gender, model misspecification, omitted variables, and measurement error. This part is termed as structural effect in this paper.

The decomposition literature has witnessed a revolution, and Fortin et al. (2011) review the decomposition methods developed since the seminal work of Oaxaca. In our study, we implement

[1] It is important to note that the decomposition in Equation (3) is performed by using coefficients obtained from the sample of male-owned firms as the reference category. Alternatively, one could also use production structure of female-owned firms as the reference category. However, the decomposition results are sensitive to the choice of reference category and leads to what is referred to as 'index number problem' in the literature (Oaxaca 1973). Many suggest using the average of male and female coefficients as the reference (Cotton 1988; Neumark 1988; Reimers 1983). Arguing that men are the usual comparison group in legal proceedings concerning gender discrimination, Ginther and Hayes (2003) recommend using coefficients from the male sample as the reference category. The standard practice in the literature on gender wage gap is also to use male coefficients as the non-discriminatory wage structure. Following this, the present study too uses coefficients of firms owned by male managers as the reference category.

three important developments in the decomposition literature. First, we estimate the productivity gap and its possible sources at various percentiles of productivity distribution (Chernozhukov et al. 2013; Firpo et al. 2007, 2009). Second, we use a reweighting approach to offset the intrinsically parametric character of the Oaxaca decomposition (Barsky et al. 2002; Firpo et al. 2007). Third, we normalize the coefficients of the categorical variables to avoid having omitted reference groups (Gardeazabal and Ugidos 2004; Yun 2005). Thus, in addition to relying on the standard Oaxaca decomposition, we perform the Oaxaca decompositions at particular quantiles in the productivity distribution using the RIF approach.

We use the RIF methodology developed by Firpo et al. (2007, 2009, 2018). The goal of the RIF approach is to move beyond the assessment of mean differences to examine the gender gap in productivity along the whole distribution of productivity using an Oaxaca-type decomposition approach based on unconditional quantile regression estimates (Davino et al. 2013; Firpo et al. 2009). In other words, this methodology helps in quantifying the role played by different variables in explaining the gap in productivity at different points in the productivity distribution. This decomposition can be performed for any distributional statistics like quantile and the Gini index (Ahmed and Maitra 2015). This procedure is implemented in two stages. In the first stage, a counterfactual distribution is created through a reweighting procedure to decompose productivity gap between male-owned and female-owned firms into an aggregate composition effect and an aggregate structural effect. The reweighting function is estimated using a logit regression.

In the second stage, the contribution of each set of explanatory variables on both of these components is ascertained. This is implemented using the RIF decomposition method, which is similar to the Oaxaca method except that the outcome variable, namely, log of labour productivity, is replaced by the RIF of the target statistic. We then estimate an ordinary least square regression of the corresponding RIF on observed firm characteristics for the male-owned firms, female-owned firms, and the counterfactual. These regression estimates are used to decompose the difference in distributional parameter between male and female entrepreneurs by replacing the log of labour productivity with the corresponding RIF for each observation and using a suitable counterfactual (Khurana and Mahajan 2020). The aggregate structural effect obtained through reweighting can be broken down into a RIF structural effect and a RIF reweighting error. In a similar vein, another decomposition can be used to decompose composition effect into a RIF composition effect and a specification error. As the RIF regressions are linear, we are able to obtain the contribution of each explanatory variable to each of the four components listed above. Our interest in this study also lies in estimating the detailed structural effect and the detailed composition effect.

# 4 Data and descriptive analysis

## 4.1 Data

Our analysis is based on the unit-level data sourced from the latest (73rd) round of the Government of India's NSSO survey on the unincorporated non-agricultural enterprises conducted in 2015–16. This nationwide enterprise-level survey has been designed to gather information on the operational and economic characteristics of the enterprises in the unincorporated non-agricultural sector. The operational characteristics covered in the survey include type of ownership, nature of operation, type of enterprise, their status of registration, constraints to their operation, government support obtained, and employment details among other attributes (NSSO 2017). The economic characteristics in the survey mainly consist of operating expenses and receipts, payments to workers, fixed assets, indebtedness, and information pertaining to loans (NSSO 2017). The survey focused on enterprises belonging to manufacturing, trade, and service sectors. These enterprises are not registered under the Factories Act of 1948, and hence are not subjected to the industrial licensing or labour laws as firms in the formal sector.[2] The survey covered all the Indian states and union territories, and used a multi-stage, stratified random sampling procedure to select the final sample of firms. Following this sample strategy, each state is divided into strata; while strata are typically districts in rural areas, cities are grouped together based on the size of population. The first-stage units (FSUs) formed the census villages in the rural sector and urban frame survey blocks in the urban sector. The enterprises, the ultimate-stage units, are selected within each FSU. In its 73rd round, the NSSO has surveyed 290,113 firms, of which 143,179 are rural firms and 146,934 are urban firms. The total sample constituted approximately 0.5 per cent of the estimated population of unincorporated non-agricultural enterprises. We therefore reweight firm-level observations using inverse sampling multipliers to make our estimates representative of the population of firms.

This NSSO dataset has been heavily used in the past by scholars to study, among other topics, growth, productivity, and survival of firms in the sector (Hsieh and Klenow 2014; Raj and Sen 2016a); dualism, missing middle, and firm transition (Kathuria et al. 2013; Kesar and Bhattacharya, 2019; Mazumdar and Sarkar 2013; Raj and Sen 2015); the impact of trade liberalization on firm performance (Nataraj 2011; Raj and Sen 2012); finance, gender, and entrepreneurship (Banerji et al. 2016; Gang et al. 2020), political reservations and female entrepreneurship (Ghani et al. 2014a); spatial determinants of entrepreneurship (Ghani et al. 2014b); and public investment and firm productivity (Chatterjee et al., 2021). However, we have not come across any study that has utilized this dataset to analyse gender-based differences in firm productivity and the factors contributing to it. The availability of crucial information on the gender of the owner enables us to address this gap in the literature.

The dataset originally consisted of 290,113 enterprises. We applied some filters to the data to generate the working sample. Our analysis was restricted to sole proprietorship firms, that is, firms with a single owner. Partnership firms, which make up about 5 per cent of the total sample, were dropped because the gender of the decision maker cannot be identified. The sample was further restricted to take care of missing observations and outliers that might bias the estimates. We omitted those firms that did not respond to one or more key questions such as age of the firm, caste of the firm owner, whether the firm maintains an account, and whether the firm undertook any work on contract basis. Finally, firms with missing values on any of the variables considered for the analysis were also excluded. These various eliminations left us with a final sample of 270,442 firms, out of which 235,566 are male-run enterprises and 34,876 are female-run enterprises.[3]

### 4.2 Construction of variables

The main objective of this paper is to check whether the gap in performance between male-owned and female-owned enterprises, as observed in majority of the existing studies, is also present in informal sector enterprises. Additionally, the paper also attempts to understand the factors contributing to this gap. Crucial, therefore, is to explain what constitutes a 'female business' and how do we construct a gender-based measure of ownership from the firm-level data. Equally

[2] A few firms in this survey are found to have employed more than 20 workers. They are certainly larger than the stipulated size to be part of the informal sector, and ideally should be counted as part of the formal sector. However, we do not know whether they are illegally operating in the sector, whether they witnessed an expansion in size after being part of the informal sector, or whether these figures are simply represent errors in data entry. Although we considered them in the final set of analyses, our results are robust to excluding them.

[3] About 7 per cent of firms in our original dataset were filtered out following these elimination norms.

important is to discuss the construction of our measure of firm performance. In this subsection, we outline the variables that we consider in the decomposition analysis and discuss the descriptive statistics (see Table 1).

Table 1: Variables and their construction

| Variables | Definition |
|---|---|
| Dependent variable | |
| Log of labour productivity | Logarithm of labour productivity, where labour productivity is defined as the ratio of gross value added to employment |
| Independent variables | |
| Firm characteristics | |
| Size of the firm | Logarithm of number of workers |
| Location of the firm | Dummy variable for firms that are located in urban areas |
| Age of the firm[a] | |
| Below 2 years | Dummy variable for firms aged less than 2 years |
| 3–9 years | Dummy variable for firms aged between 3 and 9 years |
| Above 9 years | Dummy variable for firms that have completed more than 9 years since inception |
| Any assistance from government? | Dummy variable for government assistance; it takes the value 1 if the enterprise receives any assistance from the government during the last 3 years |
| Registered under act/authority? | Dummy variable that takes the value 1 if the firm has registered under any one of the following acts: Shops and Establishment Act, Municipal Corporation/Panchayats/Local Body, Vat /Sales Tax Act, Provident Fund Act, Employees State Insurance Corporation Act, registered with SEBI/ Stock Exchange, or any other industry-specific act/authority |
| Undertake work on contract basis? | Dummy variable for firms that undertook work on contract basis |
| Accounts maintained? | Dummy variable for firms maintaining accounts |
| Firm constraints | |
| Financial constraint | A binary variable for financial constraint; the variable takes the value 1 if the firm faced any borrowing constraint in the last year, 0 otherwise |
| Electricity constraint | A binary variable for electricity constraint; the variable takes the value 1 if the firm faced any electricity constraint in the last year, 0 otherwise |
| Social group (of firm owner)[b] | |
| General category | Dummy variable for firms owned by those who belong to General category; the variable takes the value 1 if the firm is owned by an individual belonging to General category |
| Scheduled Castes (SCs) | Dummy variable for firms owned by SCs; the variable takes the value 1 if the firm is owned by an individual belonging to the SC category |
| Scheduled Tribes (STs) | Dummy variable for firms owned by STs; the variable takes the value 1 if the firm is owned by an individual belonging to the ST category |
| Other Backward Classes (OBCs) | Dummy variable for owned by OBCs; the variable takes the value 1 if the firm is owned by an individual belonging to the OBC category |

Note: SEBI, Securities and Exchange Board of India. [a]Below 2 years is the reference category for 'Age of the firm'. [b]General category is the reference category for 'Social group (of firm owner)'.

Source: authors' compilation based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Our measure of female ownership is based on a relatively straightforward definition, that is, the gender of the entrepreneur who owns and manages the enterprise. By using this definition, we create a dummy variable for gender of the owner which takes the value 1 if the firm is owned and operated by a female entrepreneur and the value 0 if it is managed by a male entrepreneur.

*Dependent variable*

The key measure that we use in the study to assess the gender differences in firm performance is productivity. We focus on a widely used partial productivity measure, namely, labour productivity. This measure is computed by dividing gross value added by total number of workers, where the workers include full-time, part-time, hired, and family workers. In the decomposition analysis, we use the logarithm of the labour productivity.

*Independent variables*

Following the discussion of determinants of firm productivity in Section 2, we group the set of explanatory variables used in our analysis into three sets: (i) *Firm characteristics*, (ii) *Firm constraints*, and (iii) *Social group*. In addition, we include two sets of controls: (iv) *Region* and (v) *Sector*. The set *Firm characteristics* includes one continuous variable and six categorical variables.[4] The continuous variable is size of the firm, which is proxied using the number of workers employed by the firm. The categorical variables include age, location, assistance, registration, linkage, and account maintenance. Age is a three-way categorical variable for age of the firm (below 2 years, 3–9 years, and above 9 years). We also distinguish between rural and urban firms through a dummy variable, location of the firm, which takes the value 1 for urban firms. Additionally, we include binary variables for firms that have received government assistance towards training and marketing, firms that have registered under any industry-specific act, firms that work solely for a contractor, and firms that maintain accounts. These variables are grouped together in the set *Firm characteristics*. In the set *Firm constraints*, we include two binary variables: one for firms that encountered any borrowing constraints and the other for firms that faced erratic power supply or power cuts. For *Social group*, we include a four-way categorical variable for social group of firm owner as a covariate of firm productivity. Based on the caste of the firm owner, the firms are classified into four categories, firms owned by General category, firms owned by SCs, firms owned by STs, and firms owned by OBCs, and grouped together under the set *Social group*.

We include two sets of controls. First, we include state dummies to control for unobserved factors at the state level that may influence firm productivity. These could be access to the coast, level of urbanization, or lack of infrastructure such as motorable roads and railway lines that limit the size of the market for informal firms, thereby inhibiting firm productivity. The set *Region* bands together dummies for 35 Indian states. The second set of controls we include are industry dummies. Firms that are in more capital-intensive industries or sectors may have higher labour productivity, for example. Firms in our data are divided into 10 different broad sectors of industrial activity. These sectoral dummies are collected together in the set *Sector*. The vector of observed covariates considered for our decomposition analysis is represented in Equation (4):

$$X = \{Firm\ Characteristics, Firm\ Constraints, Social\ group, Region, Sector\} \quad (4)$$

[4] It needs to be stated here that the RIF regression performs better when not all variables are categorical (Ferreira et al. 2021).

### 4.3 Descriptive statistics

The descriptive statistics and the results from the tests of equality between male and female entrepreneurs are presented in Table 2. This discussion will provide us with some initial insights on the gender differential in productivity and other key firm attributes. First, as expected, bulk of the firms are owned by men, with only 13 per cent of firms being owned by women. Second, we find that female-owned firms are less productive than male-owned firms. The average labour productivity of male-run firms and female-run firms are approximately 43,305 (USD 583) and 23,910 (USD 322) Indian rupees (INR), respectively. This implies that female-run firms are, on average, 45 per cent less productive than firms owned by male entrepreneurs. The male–female productivity difference of INR 19,394 (USD 261) is significant at the 1 per cent level. Third, in line with the existing evidence from other developing countries, female entrepreneurs are likely to manage younger and smaller firms than male entrepreneurs. An average firm owned by a female entrepreneur is about two-thirds the size of a firm owned by a male entrepreneur. While a male-run firm employs on average 2.5 workers (antilog of 0.3892), a female-run firm employs only 1.4 workers (antilog of 0.1663). Considering age of the firm, around 44 per cent of male entrepreneurs and 37 per cent of female entrepreneurs own the oldest firms in the sample (firms aged above 9 years). Female entrepreneurs are more likely to operate their businesses from rural areas than their male counterparts. About 56 per cent of female entrepreneurs operate their firms from rural areas while male-run firms are equally distributed between rural and urban areas. Table 2 also highlights some drastic gender differentials for government support, registration, and account maintenance. As noted in Section 2, support from the government/government agencies, registering firms under an act and maintaining regular accounts are considered important drivers of firm productivity. However, firms with these characteristics are markedly few in the entire sample. Even among the firms with these attributes, the presence of female entrepreneurs are proportionately very low. Contrarily, significantly larger shares of women entrepreneurs undertake work on contract basis. As per our estimates, about 32 per cent of female-owned firms worked for a contractor compared with a mere 5 per cent of male-owned firms.

There are also significant differences between male-run and female-run firms in terms of the constraints they face. Though firms facing these constraints are substantially few in our full sample, male entrepreneurs are more likely than female entrepreneurs to view access to finance and power availability as constraints. Firms with these constraints account for a higher share for male-owned firms: 10 per cent in financial constraint and 4 per cent in electricity constraint for male-owned firms versus 5 and 2 per cent, respectively, for female-owned firms. Finally, we see clear and sharp disparities in ownership by social group, with General category and OBCs together owning over 80 per cent of the total enterprises in the sector. This disparity in ownership by caste is also visible across male and female genders, with a marginally higher share of male entrepreneurs in General category and OBCs. Among male entrepreneurs, 51 per cent are OBCs, 33 per cent belong to General category, 12 per cent are SCs, and 4 per cent are STs. On the other hand, among female entrepreneurs 50 per cent are OBCs, 31 per cent belong to General category, 15 per cent are SCs, and 5 per cent are STs.

Table 2: Sample means

| Variables | All firms | Male-run firms | Female-run firms | Difference (2)-−(3) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Dependent variable | | | | |
| Labour productivity (in Indian rupees) | 40,803.51 | 43,304.47 | 23,910.34 | 19,394.13*** |
| | (54,245.81) | (56,525.49) | (30,154.72) | |
| Log of labour productivity | 9.9539 | 10.1662 | 9.1257 | 1.0405*** |
| | (1.0731) | (0.9686) | (1.0594) | |
| Independent variables | | | | |
| Firm characteristics | | | | |
| Size of the firm | 0.3437 | 0.3892 | 0.1663 | 0.2229*** |
| | (0.5172) | (0.5395) | (0.3688) | |
| Location of the firm | 0.4889 | 0.5009 | 0.4417 | 0.0592*** |
| | (0.4999) | (0.5000) | (0.4966) | |
| Age of the firm, below 2 years | 0.1219 | 0.1167 | 0.1422 | −0.0255*** |
| | (0.3272) | (0.3210) | (0.3493) | |
| Age of the firm, 3–9 years | 0.4516 | 0.4419 | 0.4896 | −0.0477*** |
| | (0.4977) | (0.4966) | (0.4999) | |
| Age of the firm, above 9 years | 0.4265 | 0.4414 | 0.3682 | 0.0732*** |
| | (0.4946) | (0.4966) | (0.4823) | |
| Any assistance from government? | 0.0076 | 0.0080 | 0.0057 | 0.0023*** |
| | (0.0867) | (0.0893) | (0.0755) | |
| Registered under act/authority? | 0.2953 | 0.3499 | 0.0825 | 0.2674*** |
| | (0.4562) | (0.4769) | (0.2751) | |
| Undertake work on contract basis? | 0.1032 | 0.0468 | 0.3233 | −0.2764*** |
| | (0.3042) | (0.2113) | (0.4677) | |
| Accounts maintained? | 0.0971 | 0.1129 | 0.0357 | 0.0771*** |
| | (0.2961) | (0.3164) | (0.1856) | |
| Firm constraints | | | | |
| Financial constraint | 0.0841 | 0.0943 | 0.0441 | 0.0503*** |
| | (0.2775) | (0.2923) | (0.2053) | |
| Electricity constraint | 0.0320 | 0.0349 | 0.0206 | 0.0143*** |
| | (0.1759) | (0.1835) | (0.1419) | |
| Social group (of firm owner) | | | | |
| General category | 0.3290 | 0.3343 | 0.3081 | 0.0262*** |
| | (0.4698) | (0.4717) | (0.4617) | |
| Scheduled Castes | 0.1260 | 0.1210 | 0.1456 | −0.0246*** |
| | (0.3319) | (0.3262) | (0.3527) | |
| Scheduled Tribes | 0.0413 | 0.0397 | 0.0479 | −0.0082*** |
| | (0.1991) | (0.1952) | (0.2135) | |
| Other Backward Classes | 0.5037 | 0.5050 | 0.4984 | 0.0066** |
| | (0.5000) | (0.5000) | (0.5000) | |
| Number of observations | 270,442 | 235,566 | 34,876 | |

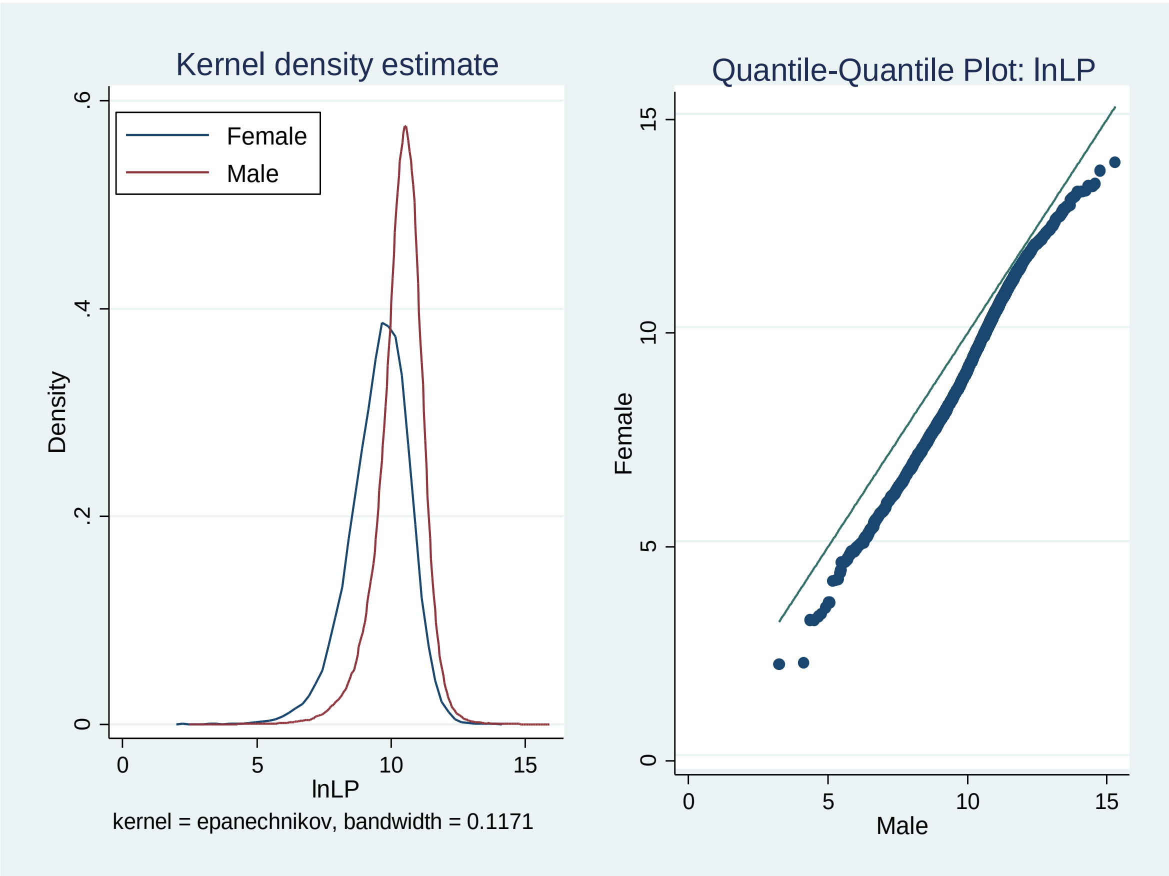
Note: standard deviations in parentheses.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

### 4.4 Productivity distributions

To further investigate the gender differences in productivity, we examine the kernel density plot and quantile–quantile plot in Figure 1. The plot on the left displays the kernel density distribution of log of labour productivity for male-owned firms and female-owned firms. The mass of the distribution of productivity for male entrepreneurs is to the right of that for female entrepreneurs, which provide evidence of gender differences in productivity. A larger gap is found on the left-tail, which continues to the middle of the distribution. However, in the right-tail of the productivity distribution, the gap in productivity between male owners and female owners nearly overlaps. The quantile–quantile plot on the right side of Figure 1 too clearly shows that male-owned firms are more productive than firms owned by female entrepreneurs. Available evidence thus points to the existence of sharp gender-related differences in firm productivity. A more detailed picture of the evolution of productivity of male-run firms and female-run firms and gender differences in productivity can be seen in Table 3, which presents the logarithm of labour productivity and the gender productivity gap at the different quantiles and at the mean. The gender gap in labour productivity at different quantiles is also presented in Figure 2.

Figure 1: Firm productivity by gender




Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

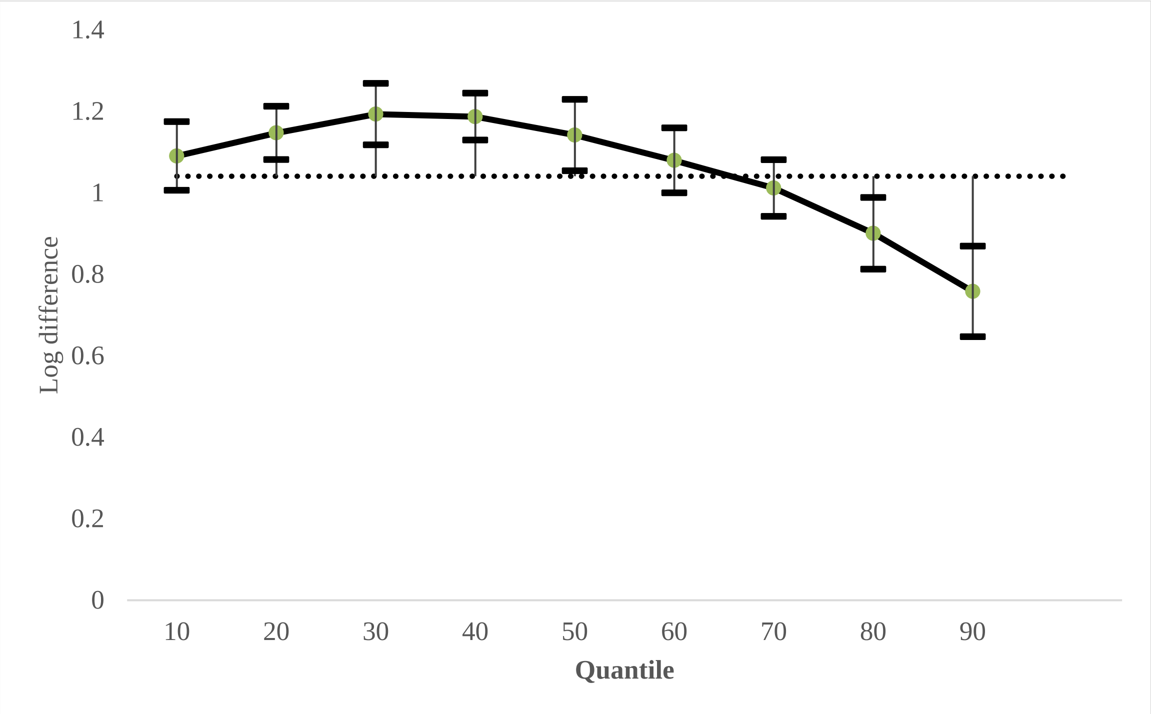
Table 3: Log labour productivity and gender productivity gap by deciles

| Percentile | Male-owned firms | Female-owned firms | Gender productivity gap |
|---|---|---|---|
| 0.10 | 10.1662 | 9.1257 | 1.0405 |
| 0.20 | 8.9267 | 7.8367 | 1.0900 |
| 0.30 | 9.4879 | 8.3412 | 1.1467 |
| 0.40 | 9.8341 | 8.6412 | 1.1929 |
| 0.50 | 10.0834 | 8.8968 | 1.1866 |
| 0.60 | 10.2867 | 9.1454 | 1.1414 |
| 0.70 | 10.4887 | 9.4095 | 1.0793 |
| 0.80 | 10.6893 | 9.6778 | 1.0115 |
| 0.90 | 10.9173 | 10.0169 | 0.9003 |
| Mean | 11.2318 | 10.474 | 0.7578 |

Note: the productivity gap is the difference between the log labour productivity of male-run and female-run firms. Weights used in the calculations.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Figure 2: Gender productivity gap at quantiles




Note: log difference is between male and female labour productivity. The dashed line shows the gender gap at the mean. Entries are based on the reweighted RIF-Oaxaca decomposition results presented in Table 5.

Source: authors' elaboration based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

It is clearly evident from Table 3 that the productivity of male-owned firms is higher than that of female-owned firms at all percentiles. When we look at the gender productivity gap, which is defined as log productivity of male-run firms minus log productivity of female-run firms, we find that the mean gender productivity gap is 70 log points. Comparing productivity at different percentiles of the gender-specific productivity distributions, we find that gender productivity gap takes a mild inverted U-shape until about the 70th percentile before declining rapidly at the top end of the productivity distribution (Table 3, last column, and Figure 2). We find that the gap is strictly increasing up to the 40th percentile and thereafter falls steadily. After reaching a peak of 119 log points at the 40th percentile, it drops to 90 log points at the top percentile. In other words, our gender-specific productivity distribution shows that the gender gap in productivity is largest at the bottom percentiles and lowest at the top percentiles, which corroborates our conjectures based on kernel density plot. The gender productivity varying across the productivity distribution implies that focusing on the mean productivity gap per se may not be informative. In our empirical strategy, we attempt to examine the heterogeneity we observe in the gender productivity gap by using the RIF decomposition method that is especially suited for this purpose.

## 5 Empirical analysis and discussion

### 5.1 Baseline determinants of productivity of male-owned and female-owned firms

Table 4 presents the ordinary least square estimates of a regression of the logarithm of labour productivity on the set of explanatory variables reported in Table 2. Importantly, it compares the effects of various factors across genders. While endogeneity may be an issue here, our objective is not to infer causality but to assess the importance of these factors, especially gender, in explaining labour productivity.

Table 4: Determinants of log productivity, ordinary least square estimates

| | Male-run firm | Female-run firm |
|---|---|---|
| Intercept | 9.1157*** | 9.2454*** |
| | (0.2146) | (0.3057) |
| Firm characteristics | | |
| Size of the firm | −0.2642*** | −0.2857*** |
| | (0.0225) | (0.0489) |
| Location of the firm | 0.4027*** | 0.2360*** |
| | (0.0238) | (0.0397) |
| Age of the firm, 3–9 years | 0.4634*** | 0.7253*** |
| | (0.0257) | (0.0524) |
| Age of the firm, above 9 years | 0.4327*** | 0.7506*** |
| | (0.0279) | (0.0506) |
| Any assistance from government? | 0.2117*** | 0.0947 |
| | (0.0496) | (0.1151) |
| Registered under act/authority? | 0.4087*** | 0.6624*** |
| | (0.0192) | (0.0463) |
| Undertake work on contract basis? | −0.1281** | −0.1603** |
| | (0.0566) | (0.0654) |
| Accounts maintained? | 0.3947*** | 0.5270*** |
| | (0.0244) | (0.0730) |
| Firm constraints | | |
| Financial constraint | −0.0470 | −0.1076** |
| | (0.0319) | (0.0532) |
| Electricity constraint | 0.1073*** | 0.0454 |
| | (0.0324) | (0.0836) |
| Social group (of firm owner) | | |
| Scheduled Castes | −0.2627*** | −0.1842*** |
| | (0.0263) | (0.0484) |
| Scheduled Tribes | −0.6008*** | −0.3430*** |

| | | |
|---|---|---|
| | (0.0865) | (0.1127) |
| Other Backward Classes | −0.1373*** | −0.1000*** |
| | (0.0202) | (0.0392) |
| Regional effects? | Yes | Yes |
| Sectoral effects? | Yes | Yes |
| R-squared | 0.2929 | 0.3015 |
| Number of observations | 235566 | 34876 |

Note: the dependent variable is the logarithm of labour productivity. Standard errors reported in parentheses are robust to heteroskedasticity and clustered residuals within districts. Sampling weights are used in estimations. ***, **, and * denote significance at the 1, 5, and 10 per cent levels, respectively.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Strikingly, we find across the two genders that labour productivity declines with firm size. Smaller firms are more productive than larger ones, and the magnitude does not vary significantly across gender. Location of the firm is another significant factor influencing firm productivity. In line with the evidence available in the literature, urban firms are more productive than rural firms, regardless of gender. The magnitude of impact, however, is higher for male-owned firms than for female-owned firms: urban male-owned firms are 40 per cent more productive than rural male-owned firms while the corresponding number for female-owned firms is 24 per cent. Age of the firm appears to have a significant influence on labour productivity. The binary variables representing two firm age categories, firms aged between 3 and 9 years and above 9 years, yield positive and significant coefficients, indicating that older firms are likely to be more productive than younger firms. The finding is in line with the extant evidence from other developing countries. This age–productivity nexus is consistent across genders but, as evident from the coefficient values, the effect of age on productivity is substantially larger for firms owned by women. However, this result might suggest that female entrepreneurs are further at a disadvantage as they own younger firms, on average (Table 2). We also find that productivity is higher for firms that are registered, firms that received government support, and firms that maintained accounts, in line with our priors based on the discussion in Section 2. On the other hand, our results reveal that firms that work for a contractor seem to suffer productivity losses compared with firms that do not work for a contractor. In other words, firms that have entered into subcontracting arrangements are less productive than firms that have not, and this finding holds regardless of the gender of the owner.

Our measures of firm constraints seem to have a differential effect across genders. For financial constraints, we find a negative effect on productivity for male owners but no effect for female owners. Surprisingly, the variable proxying electricity constraints yields a positive coefficient for male owners and an insignificant coefficient for female owners. Our results clearly suggest that social group of the firm owner is an important determinant of firm productivity. Irrespective of the gender of the owner, the binary variables for SCs, STs, and OBCs return a negative coefficient, suggesting that productivity of firms with SC, ST, and OBC owners is lower than that of firms with owners from General category. This finding corroborates the existing evidence from other studies that there are social and economic barriers to informal sector enterprises in increasing their productivity.

### 5.2 Oaxaca mean decomposition of the sources of gender productivity gaps

The descriptive analyses in the preceding sections have confirmed the existence of gender productivity gaps. However, it is important to identify the factors that contribute to the widening of these gaps, which help us to suggest measures and interventions likely to reduce or even close the gaps. To unpack the potential sources of gender gap, we decompose the gender productivity differential into components describing the contribution of observed characteristics by using the Oaxaca decomposition method.

We now discuss the results of this empirical exercise. We start with the standard Oaxaca decomposition results, and then discuss the RIF-Oaxaca decompositions at the mean and different deciles. In all our decompositions, the variables used in our model are grouped into different covariate sets, as discussed in the previous section, to facilitate an interpretation of the results. We formed five such variable sets: (i) *Firm characteristics*, (ii) *Firm constraints*, (iii) *Social group*, (iv) *Region*, and (v) *Sector*. The contribution of the variable set is simply the sum of the contributions of each variable included in the set. The results from the basic Oaxaca decomposition are presented in Table 5. The table presents the results of the decomposition of the gender gap in log labour productivity into composition and structural effects, both overall and for each (set of) covariate(s). Both coefficient estimates[5] and the percentage contribution of each covariate to the aggregate effect are presented. For interpretation, we focus on the percentage share that tells us the proportion of the gender productivity gap attributable to differences in observed attributes and what proportion to differences in the returns to these observables.

Table 5: Standard Oaxaca decomposition of gender productivity gap

| | Composition effect | | Structural effect | |
|---|---|---|---|---|
| | Estimate | Share (%) | Estimate | Share (%) |
| Aggregate effects | 0.2819*** | 27.09 | 0.7586*** | 72.91 |
| | (0.0359) | | (0.0405) | |
| Intercept | N/A | N/A | 0.4447*** | 58.62 |
| | | | (0.1087) | |
| Firm characteristics | 0.1502*** | 53.28 | 0.0456 | 6.01 |
| | (0.0207) | | (0.0770) | |
| Firm constraints | −0.0008 | −0.28 | −0.0573 | −7.55 |
| | (0.0017) | | (0.0470) | |
| Social group | 0.0105* | 3.72 | 0.0510* | 6.72 |
| | (0.0059) | | (0.0260) | |
| Regional effects | 0.0222 | 7.88 | 0.1826*** | 24.07 |
| | (0.0179) | | (0.0337) | |
| Sectoral effects | 0.0998*** | 35.40 | 0.0920 | 12.13 |
| | (0.0124) | | (0.0631) | |

Note: N/A, not applicable. The dependent variable is the logarithm of labour productivity. Robust standard errors in parentheses are clustered at the district level. Sampling weights are used in estimations. ***, **, and * denote significance at the 1, 5, and 10 per cent levels, respectively. Each category summarizes the contribution of the sum of individual effects. 'Firm characteristics' summarizes the contribution of size, location, age, government assistance, registration, subcontracting, and account maintenance. 'Firm constraints' captures the combined effect of financial constraint and electricity constraint. 'Social group' summarizes the effect of dummies representing the social group of the firm owner. 'Regional effects' adds dummies for Indian states. 'Sectoral effects' bands together dummies for broad sectors. Share is computed as a proportion to the predicted gender productivity gap.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Both composition and structural effects carry positive signs and are significant at the 1 per cent level. Results show that structural effect is far more important than composition effect in explaining the gap in productivity between male-owned and female-owned firms, as revealed by the size of the coefficients. Composition effect, that is, the proportion of the gender productivity

[5] Positive coefficient widens the gap while a negative coefficient reduces the gap.

gap due to the difference in the levels of observables between male and female entrepreneurs, explains almost a quarter (27 per cent) of the gender gap whereas structural effect, representing the portion of gender differential attributable to return to the same observed factors, accounts for 73 per cent of the gap. This means that male entrepreneurs, who have better endowments, benefit more from the observed factors than their female counterparts. In addition, it can be clearly discerned that the male owners have a clear structural advantage in terms of the returns to observable characteristics. It is primarily female owners with structural disadvantages in the returns to observed characteristics that are driving their lower productivity levels, according to our study. Indeed, our results suggest that if the coefficients of the variables influencing firm productivity yielded similar returns for both male-owned and female-owned firms, then 73 per cent of the gap in productivity between them would be reduced. On the other hand, the gap in productivity would have seen a drop by 27 per cent if the female-owned firms had similar endowments and advantages in characteristics as male-owned firms.

We also break down composition and structural effects into important covariate sets in Table 5. In interpreting the results, a positive coefficient of a covariate tends to widen the gender productivity gap, while a negative coefficient of a covariate narrows the gap. The results of the detailed decomposition show that firm characteristics, social group of the firm owner, and sectoral effects contribute significantly to the size of the composition effect. These factors amplify the gender gap by augmenting the composition effect. The biggest contribution to the productivity gap via composition effect originates from the male–female differential in firm characteristics. A total of 15 percentage points of the gender gap can be attributed to the difference in the characteristics of firms owned by men and women. In relative terms, more than half of the explained gap (53.28 per cent) is explained by the gender difference in firm characteristics. This suggests that if female-owned firms possessed the same levels of characteristics as male-owned firms, then the contribution to the productivity gap by differences in observed characteristics would have declined by 53 per cent. The significant gender differential in firm characteristics is already observed in Table 2. Firm attributes as an important predictor of gender gap in firm performance has been already documented in the literature. Studies have highlighted the significant role of firm size (Marlow and McAdam 2013), firm age (Chaudhuri et al. 2020), and firm location (Raj and Sen 2016b) in explaining the gender differences in firm performance. Aterido et al. (2011) find that one-third of the productivity gap is explained by the differences in the types of enterprises women own: smaller firms, firms that are unaffiliated with other businesses, and firms that are not registered.

Our results also show the importance of composition effect for gender differences in sectoral choice in determining the productivity gap, as captured by the estimates on sectoral effects. Gender differences in sectoral choice contribute about 10 percentage points to the gender gap, which corresponds to roughly 36 per cent of the gap explained by composition effect. This implies that one-third of the gap can be explained by differential sorting of male-owned and female-owned firms into sectors or industries. The male advantage in the industry variables indicates that many female-owned firms are located in less productive sectors as opposed to male-run firms, which eventually widens the productivity gap. The intensification of gender productivity gap owing to sectoral choice is consistent with the existing evidence in the literature for developing countries. It is argued that the sectors in which women tend to operate are more crowded and register lower profits and growth potential than male-dominated sectors (Carranza et al. 2018). The predominance of female entrepreneurs in low-performing and less productive industries tends to explain the underperformance of the businesses they operate (Bardasi et al. 2011). De Mel et al. (2009) maintain that once differences in sector are accounted for, there is no longer a significant difference between male-owned and female-owned enterprises.[6] In the Indian context too, there is evidence that female entrepreneurs are more likely to be in sectors that are dominated by less productive firms (Chaudhuri et al. 2020).

The third important factor that contributes to composition effect of the gender gap is gender differences in the social group of the firm owner. Compared with firm characteristics and sectoral choice, the social group of the firm owner has a minor role to play in explaining the gender gap in productivity between male-owned and female-owned enterprises; just over 1 percentage point of the gap can be attributed to gender differences in social group, which amounts to a contribution of about 4 per cent in relative terms. There is enough evidence to show that firms owned by members from disadvantaged groups such as SCs and STs are less productive than those headed by members of the General category (Raj and Sen 2016b). More importantly, the share of female-owned enterprises is significantly higher among SC- and ST-owned enterprises (Deshpande and Sharma 2013). In their study on vegetable sellers in India, Delecourt and Ng (2021) observed that female entrepreneurs disproportionately come from disadvantaged backgrounds.

Having seen that the bulk of the gender gap in productivity is explained by structural effect, it would be interesting to see which factors contributed to the augmentation of the productivity gap by enlarging the structural effect. The structural effect is almost entirely explained by unidentified firm attributes. This is evident from the coefficient of the intercept term, which is positive, large, and significant at the 1 per cent level. The gender difference in the returns to unobserved factors contribute to about 45 percentage points to the observed productivity gap, nearly three-fifths of the total contribution to the productivity gap by structural effect. Of all the explanatory variables included, social group of the firm owner and regional effects have yielded a significant coefficient. The most important structural effect was the increasing productivity gaps associated with spatial location of firms. On its own, it contributed 18.26 percentage points to the increase in productivity gap, accounting for 24 per cent of the gap explained by structural effect. The regional effects point to the distribution of firms across regions. The substantial advantage of male-run firms in the regional dummies indicates that male-run firms are more likely to be in states with higher average productivity than female-run firms. In particular, male-run firms are more likely to locate in regions where firms are highly productive, which increases their productivity indirectly. But although it was the most important factor, structural effect of spatial location of firms did not act alone. Gender differences in the coefficients of social group also have significant positive effects on the gender productivity gap. This may indicate that female-run firms owned by socially disadvantaged groups have a double disadvantage, where caste and gender interact to accentuate the gender gap in productivity. Finally, the fact that the differences in firm constraints do not explain any of the components of Oaxaca decomposition is indicative of their importance for only a small percentage of firms.

### 5.3 RIF-Oaxaca distributional decomposition of the sources of gender productivity gaps

We now turn to the results from the RIF-based decomposition. As mentioned earlier, RIF-Oaxaca decomposition uses RIF and the traditional Oaxaca method in conjunction with reweighting to decompose the gender gap in productivity across overall distributions into composition and structural effects. Further, as is the case with standard Oaxaca decomposition, it also returns the contribution of each explanatory variable on both aggregate effects. Results are presented in the next two subsections. The first subsection presents the decompositions at the average labour productivity. The second subsection examines whether the gender gap in productivity and its

[6] A study by Campos and Gassier (2015) for Uganda show that the women who crossed over to male-dominated sectors reported higher returns than those that remained in female-dominated sectors.

determinants differ at different points in the distribution of labour productivity. The gender gap is evaluated at every decile of the labour productivity distribution. For each case, besides decomposing the gender gap in productivity into structural and composition effects, detailed decompositions of each effect by individual covariates are also presented.

### 5.4 Decomposition of average productivity gap

Table 6 presents the results for the RIF-Oaxaca gender composition of differences in average labour productivity. The first column shows the estimates of mean decomposition and the second column reports the percentage share that tells us what percentage of the gender productivity gap is accounted for by a particular covariate or a group of covariates. Counterfactual stands for estimated productivity distribution showing what would have been the labour productivity of female-run firms if they had similar coefficients as male-run firms. The pure components are the differences net of specification and reweight errors.

One potential concern related to the Oaxaca decomposition is whether the linearity assumption is satisfied, which is very much important for the consistent estimation of composition and structural effects (Firpo et al. 2018). Results in Table 6 show that the decomposition of the gender gap at the mean yields a specification error of 0.004, which is not statistically different from zero. This implies that the linear specification is justified empirically, and any apprehension related to the misspecification of the model can be ruled out. Moreover, the reweighting error is very close to zero and statistically insignificant, indicating that reweighting factors are consistently estimated.

The decomposition results based on RIF-Oaxaca are not substantially different from the ones arrived at by using the traditional Oaxaca decomposition method. Structural effect accounts for the bulk of the gender gap in productivity. Composition effect contributes to one-fourth of the gender gap, with major contributions originating from gender differences in firm characteristics, social group, and sectors of choice. Of the pure explained gender gap in labour productivity, differences in firm characteristics, social group, and sectoral choice explain approximately 52, 5, and 35 per cent, respectively. The only noticeable difference is that firm characteristics, social group, and regional effects lose their significance in the pure structural effect.

Table 6: RIF-Oaxaca decomposition at the mean

| | (1) | (2) |
|---|---|---|
| | Estimate | Share (%) |
| Overall | | |
| Mean male labour productivity (M) | 10.1662*** (0.0289) | |
| Mean female labour productivity (F) | 9.1257*** (0.0666) | |
| Gender gap in productivity (M−F) | 1.0405*** (0.0557) | |
| Reweighting decomposition | | |
| Counterfactual (C) | 9.8973*** (0.0338) | |
| Total composition effect (M−C) | 0.2689*** (0.0287) | 25.84 |
| Total structural effect (C−F) | 0.7716*** (0.1216) | 74.16 |
| RIF aggregate decomposition | | |
| Pure composition effect | 0.2643*** (0.0429) | 98.29 |
| Specification error | 0.0046 (0.0481) | 1.71 |
| Pure structural effect | 0.7581*** (0.1187) | 98.25 |
| Reweighting error | 0.0135 (0.0246) | 1.75 |
| Pure composition effect | | |
| Firm characteristics | 0.1380*** (0.0237) | 52.21 |
| Firm constraints | −0.0002 (0.0022) | −0.08 |
| Social group | 0.0137** (0.0060) | 5.18 |

| | | |
|---|---|---|
| Regional effects | 0.0197 (0.0220) | 7.45 |
| Sectoral effects | 0.0931*** (0.0175) | 35.23 |
| Pure structural effect | | |
| Firm characteristics | 0.0682 (0.2934) | 9.00 |
| Firm constraints | −0.0248 (0.1825) | -3.27 |
| Social group | 0.0586 (0.0944) | 7.73 |
| Regional effects | 0.1289 (0.1315) | 17.00 |
| Sectoral effects | 0.1618 (0.2424) | 21.34 |
| Intercept | 0.3652 (0.4058) | 48.17 |
| Observations | 270,442 | |

Note: the dependent variable is the logarithm of labour productivity. Standard errors reported in parentheses are robust to heteroskedasticity and clustered residuals within districts. Sampling weights are used in estimations. The reweighting factors are estimated using a logit model. ***, **, and * denote significance at the 1, 5, and 10 per cent levels, respectively. Counterfactual (C) is the estimated distribution of productivity, showing female mean productivity (or productivity gap) if they had the same coefficients as their male counterparts. Total composition effect refers to the part of the productivity gap due to gender differences in characteristics/endowments. Total structural effect refers to the part of the productivity gap due to gender differences in returns to those characteristics. The pure composition effect and pure structural effect are the differences net of specification error and reweighting error, respectively. Each category summarizes the contribution of the sum of individual effects. ‘Firm characteristics’ summarizes the contribution of size, location, age, government assistance, registration, subcontracting, and account maintenance. ‘Firm constraints’ captures the combined effect of financial constraint and electricity constraint. ‘Social group’ summarizes the effect of dummies representing the social group of the firm owner. ‘Regional effects’ adds dummies for Indian states. ‘Sectoral effects’ bands together dummies for broad sectors. Share is computed as a proportion of the predicted gender productivity gap.

Source: authors’ calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

### 5.5 Decomposition at unconditional quantiles

To gain further insights into how the gender productivity gap evolves along the productivity distribution, we evaluate the gender gap at every decile of the productivity distribution. It also permits us to estimate the contribution of each covariate in explaining productivity gap at different productivity quantiles, as part of either composition effect or structural effect. We provide the gender gap estimates and the aggregate RIF decompositions at each decile of the productivity distribution in Table 7. The graphical representation of these findings are reported in Figure 3.

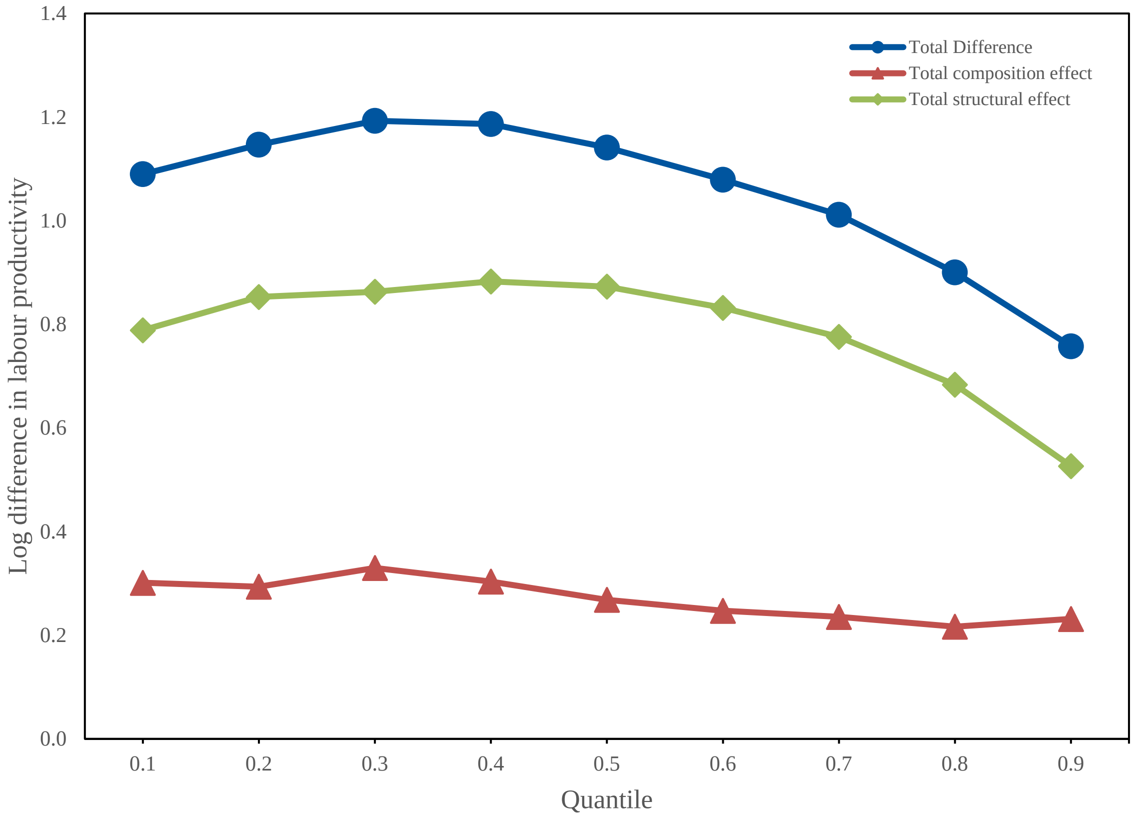
Table 7: Decomposing productivity gap by percentiles: RIF-Oaxaca aggregate decomposition

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| Overall | | | | | | | | | |
| Mean male labour productivity (M) | 8.9267*** | 9.4879*** | 9.8341*** | 10.0834*** | 10.2867*** | 10.4887*** | 10.6893*** | 10.9173*** | 11.2318*** |
| | (0.0462) | (0.0356) | (0.0307) | (0.0268) | (0.0262) | (0.0264) | (0.0257) | (0.0282) | (0.0279) |
| Mean female labour productivity (F) | 7.8367*** | 8.3412*** | 8.6412*** | 8.8968*** | 9.1454*** | 9.4095*** | 9.6778*** | 10.0169*** | 10.4740*** |
| | (0.0976) | (0.0751) | (0.0829) | (0.0639) | (0.0949) | (0.0888) | (0.0788) | (0.0984) | (0.1217) |
| Gender gap in productivity (M−F) | 1.0900*** | 1.1467*** | 1.1929*** | 1.1866*** | 1.1414*** | 1.0793*** | 1.0115*** | 0.9003*** | 0.7578*** |
| | (0.0842) | (0.0655) | (0.0755) | (0.0576) | (0.0875) | (0.0797) | (0.0694) | (0.0881) | (0.1112) |
| Reweighting decomposition | | | | | | | | | |
| Counterfactual (C) | 8.6254*** | 9.1941*** | 9.5043*** | 9.7797*** | 10.0183*** | 10.2414*** | 10.4537*** | 10.7006*** | 11.0003*** |
| | (0.0558) | (0.0378) | (0.0408) | (0.0413) | (0.0374) | (0.0322) | (0.0341) | (0.0307) | (0.0299) |
| Total composition effect (M−C) | 0.3013*** | 0.2938*** | 0.3299*** | 0.3037*** | 0.2685*** | 0.2473*** | 0.2356*** | 0.2166*** | 0.2315*** |
| | (0.0530) | (0.0382) | (0.0384) | (0.0351) | (0.0301) | (0.0264) | (0.0277) | (0.0262) | (0.0258) |
| Total structural effect (C−F) | 0.7887*** | 0.8529*** | 0.8630*** | 0.8829*** | 0.8729*** | 0.8319*** | 0.7759*** | 0.6837*** | 0.5263** |
| | (0.2202) | (0.1658) | (0.1735) | (0.1146) | (0.1862) | (0.1635) | (0.1469) | (0.1970) | (0.2580) |
| RIF aggregate decomposition | | | | | | | | | |
| Pure composition effect | 0.3702*** | 0.3554*** | 0.3103*** | 0.2636*** | 0.2325*** | 0.2133*** | 0.1905*** | 0.1995*** | 0.2059*** |
| | (0.1375) | (0.0710) | (0.0490) | (0.0339) | (0.0321) | (0.0262) | (0.0236) | (0.0218) | (0.0273) |
| Specification error | −0.0689 | −0.0616 | 0.0196 | 0.0401 | 0.0359 | 0.0340 | 0.0451* | 0.0171 | 0.0256 |
| | (0.1525) | (0.0800) | (0.0557) | (0.0364) | (0.0331) | (0.0281) | (0.0242) | (0.0233) | (0.0257) |
| Pure structural effect | 0.7921*** | 0.8499*** | 0.8461*** | 0.8675*** | 0.8559*** | 0.8096*** | 0.7483*** | 0.6585*** | 0.5002* |
| | (0.2213) | (0.1670) | (0.1720) | (0.1117) | (0.1821) | (0.1597) | (0.1427) | (0.1931) | (0.2553) |
| Reweighting error | −0.0034 | 0.0030 | 0.0169 | 0.0154 | 0.0170 | 0.0223 | 0.0275 | 0.0251 | 0.0261 |
| | (0.0341) | (0.0262) | (0.0327) | (0.0322) | (0.0298) | (0.0261) | (0.0267) | (0.0226) | (0.0211) |

Note: the dependent variable is the estimated RIF at the respective decile of the log labour productivity. The gender gap is the difference between the log labour productivity of male-owned and female-owned enterprises. Sampling weights are used in estimations. Standard errors reported in parentheses are robust to heteroskedasticity and clustered residuals within districts. The reweighting factors are estimated using a logit model. ***, **, and * denote significance at the 1, 5, and 10 per cent levels, respectively. See Table 5 for details of the variables included in each category.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Figure 3: Breakdown of productivity gap by deciles




Note: log difference is between male and female labour productivity. Entries are based on the reweighted RIF-Oaxaca decomposition results presented in Table 5.

Source: authors' elaboration based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Our results unambiguously highlight the importance of investigating the gender productivity gap and its causes across the full productivity distribution. The average gender productivity gap conceals important subtleties along the productivity distribution. First, the gender gap is distributed unevenly across the productivity distribution. Further, the estimated productivity of male-owned and female-owned firms increase as we move from the bottom to the top deciles, though at different rates (Table 7). Another interesting point emerging from Table 7 is that the estimated gender productivity gaps are positive and significant at each decile. But the productivity differential is most pronounced at the bottom and middle deciles and the difference narrows considerably at the top of the distribution. The gender gap at the 90th percentile is the lowest, which is 28 per cent less than the mean gap of 1.0405 reported in Table 6. Instead, the gender gap at the 10th percentile through the 60th percentile lies above the average gender gap in productivity. In other words, gender productivity differentials clearly shrink as we move up the productivity distribution.[7] The key finding emerging from this analysis is that the gender differences are substantially large among less productive firms and relatively small among highly productive firms.

[7] This is also confirmed by our results for the standard measures of top-end (90–50 log productivity differential) and low-end (50–10 log productivity differential) gender gap in productivity, as well as for the variance of log labour productivity and the Gini coefficient (see Appendix Table A2). The gender gap in productivity for 90–10 and 90–50 differences are negative and significant, suggesting that the gender gap tends to narrow as we move up the productivity distribution.

Even if we assume heterogeneity of firms along the productivity distribution, female-run firms seem to be always disadvantaged.

As observed for the average productivity, the productivity gap between male-owned and female-owned firms is primarily driven by structural effect (Figure 3). However, its contribution to gender gap varies considerably when we look at different points of the productivity distribution. In fact, the explanatory power of structural effect is the highest between the 20th and the 60th percentiles, where its contribution ranges from 83 to 85 percentage points. Conversely, the contribution of structural effect amounts to only 53 percentage points at the 90th percentile (Table 7).[8] Although its contribution varies across the deciles in absolute terms, its relative importance does not vary substantially (Table 8). In relative terms, its contribution ranges between 70 and 77 per cent. When we look at the contribution of composition effect, we find that the gender differences in observed factors are clearly in favour of male-owned enterprises along the productivity distribution.

Table 8: RIF-Oaxaca decomposition by percentiles: percentage contribution

| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|
| Gender gap | 1.09 | 1.1467 | 1.1929 | 1.1866 | 1.1414 | 1.0793 | 1.0115 | 0.9003 | 0.7578 |
| | | | | | | | | | |
| Composition (%) | 27.64 | 25.62 | 27.66 | 25.59 | 23.52 | 22.91 | 23.29 | 24.06 | 30.55 |
| Firm characteristics (%) | 18.75 | 16.73 | 14.06 | 11.54 | 10.88 | 10.32 | 10.21 | 11.11 | 11.89 |
| Firm constraints (%) | 0.28 | 0.06 | 0.01 | −0.07 | −0.10 | −0.07 | −0.07 | −0.04 | −0.15 |
| Social group (%) | 2.56 | 1.55 | 1.10 | 0.84 | 0.68 | 0.71 | 0.66 | 0.68 | 0.71 |
| Regional effects (%) | 1.86 | 1.83 | 1.30 | 1.57 | 1.45 | 1.89 | 1.90 | 2.59 | 4.25 |
| Sectoral effects (%) | 10.51 | 10.81 | 9.56 | 8.33 | 7.46 | 6.91 | 6.14 | 7.84 | 10.48 |
| Specification error (%) | −6.32 | −5.37 | 1.64 | 3.38 | 3.15 | 3.15 | 4.46 | 1.90 | 3.38 |
| | | | | | | | | | |
| Structure (%) | 72.36 | 74.38 | 72.34 | 74.41 | 76.48 | 77.08 | 76.71 | 75.94 | 69.45 |
| Firm characteristics (%) | 10.13 | 18.23 | 16.23 | 17.98 | −1.01 | 5.05 | 9.13 | 18.22 | 69.42 |
| Firm constraints (%) | 14.47 | −9.35 | −3.74 | 0.73 | 5.15 | 7.20 | 7.33 | 9.20 | 8.55 |
| Social group (%) | 18.42 | 9.07 | 3.93 | 6.46 | 2.51 | 2.52 | 2.22 | 3.15 | −4.49 |
| Regional effects (%) | 1.02 | −2.46 | 4.46 | 3.08 | 15.01 | 19.38 | 11.63 | 28.02 | 44.43 |
| Sectoral effects (%) | 5.77 | 24.38 | 20.97 | 14.66 | 17.47 | 21.42 | 26.10 | 7.09 | 8.06 |
| Reweighting error (%) | −0.31 | 0.26 | 1.42 | 1.30 | 1.49 | 2.07 | 2.72 | 2.79 | 3.44 |
| Intercept (%) | 22.85 | 70.70 | 61.55 | 66.16 | 35.86 | 19.43 | 17.58 | 7.46 | −59.98 |

Note: the dependent variable is the estimated RIF at the respective decile of the log labour productivity. The gender gap is the difference between the log labour productivity of male-owned and female-owned enterprises. Sampling weights are used in estimations. Each category summarizes the contribution of the sum of individual effects. 'Firm characteristics' summarizes the contribution of size, location, age, government assistance, registration, subcontracting, and account maintenance. 'Firm constraints' captures the combined effect of financial constraint and electricity constraint. 'Social group' summarizes the effect of dummies representing the social group of the firm owner. 'Regional effects' adds dummies for Indian states. 'Sectoral effects' bands together dummies for broad sectors. Share is computed as a proportion to the predicted gender productivity gap.

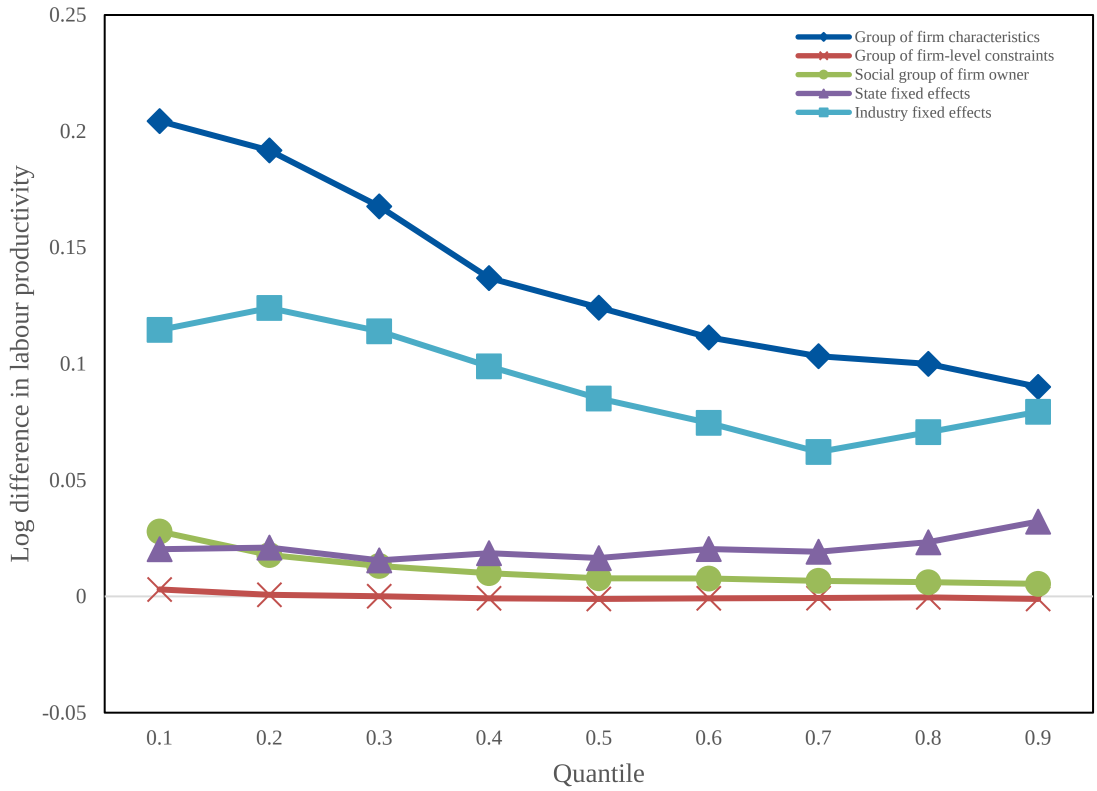
Source: authors' elaboration based on the results presented in Appendix Table A1.

But which are the most important covariates explaining the gender differences in firm productivity? Does their importance vary across the productivity distribution? To address these questions, we look at the detailed decomposition results. The RIF regressions underlying the detailed compositions are reported in Appendix Table A1. The graphical representation of these findings are reported in Figures 4 and 5. The important covariates that will be the focus of our

[8] This is confirmed by Appendix Table A2 where the coefficient of structural effect, though insignificant, carries a negative sign and has a large coefficient for top-end productivity differential.

discussion based on the RIF detailed decomposition results are in line with those that have been emphasized as part of the mean decomposition results. Their relative importance, however, changes across the distribution.

Figure 4: Detailed decomposition of composition effect




Note: log difference is between male and female labour productivity. Entries are based on the reweighted RIF-Oaxaca decomposition results presented in Table 5.

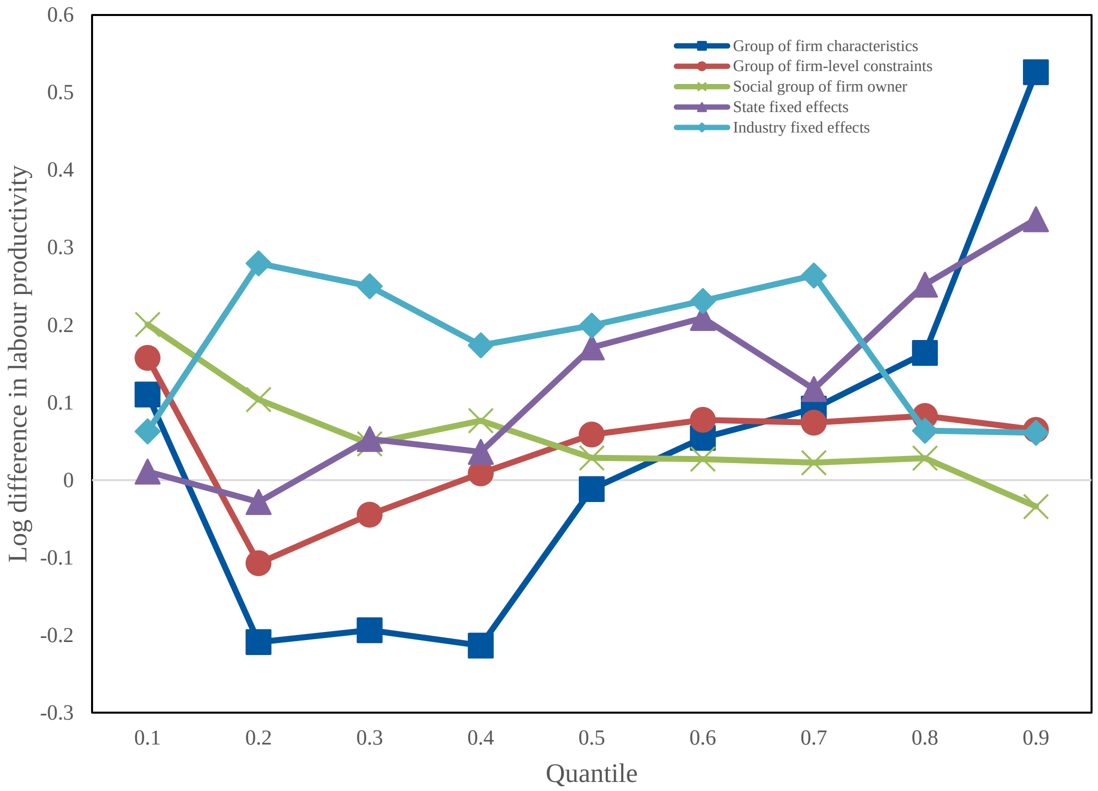
Source: authors' elaboration based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Figure 4 and Appendix Table A1 clearly show that the lion's share of the productivity gap between male-owned and female-owned firms via composition effect is explained by firm characteristics. The contribution of firm characteristics is positive and significantly different from zero at every decile. However, we also see a reduction in its contribution as we move along the productivity distribution.[9] While about 19 per cent of the productivity gap roots in differences in firm characteristics in terms of size, age, location, assistance, registration, linkage, and account maintenance at the bottom decile, its contribution reduces to almost 10 per cent at the top decile. Although the magnitude of the coefficient of firm characteristics decreases in absolute and relative terms, it remains economically significant and indicates the importance of bridging the differences in firm attributes to lessen the gap in labour productivity for all firms. The covariate on sectoral effects too is associated with sustained, positive contributions towards composition effect at each decile of interest. As observed for firm characteristics, the contribution of sectoral effects also consistently declines as we move from the bottom to the top of the productivity distribution. Depending on the decile, the portion of the gap that can be assigned to gender differences in

[9] This is corroborated by the coefficient of firm characteristics in Appendix Table A2. The coefficient yields a negative sign, though insignificant.

sectoral choice ranges between 6 and 11 percentage points, with higher values at the bottom of the distribution. The gender gap widening effect of social group of the firm owner seems to be more pronounced at the first half of the productivity distribution.

Figure 5: Detailed decomposition of structural effect




Note: log difference is between male and female labour productivity. Entries are based on the reweighted RIF Oaxaca decomposition results presented in Table 5.

Source: authors' elaboration based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

## 6 Conclusion

Using a large nationally representative dataset for the informal sector in India, we examine the patterns and correlates of gender inequality in the productivity differences between male-owned and female-owned firms. We find systematic differences in productivity between male-owned and female-owned firms, with male-owned firms more productive than female-owned firms. However, this gender gap in productivity is particularly observed in the bottom and middle parts of the productivity distribution, and less evident for the most productive male-owned and female-owned firms.

To understand the gender differences in productivity across the entire distribution, we use the Oaxaca and RIF decomposition methods to separate the role of differences in observable factors such as firm size and age, social group of the firm owner, and locational factors (composition effect) from differences in returns to these observable factors (structural effect). We find that about 73 per cent of the productivity gap can be explained by structural effect and the remaining gap explained by differences in observable characteristics (composition effect). We also find that among observable characteristics, the most important contributing factors that explain the gender productivity gap are firm characteristics for both composition and structural effects. Male-owned firms are more advantaged in firm characteristics such as firm size, age of the firm, assistance from the government, registration with state authorities, working on a contract basis, and maintaining accounts as well as earning more for the same characteristics. Regional and sectoral effects also matter in explaining structural effect and the social group of the firm owner and regional factor in explaining composition effect. Interestingly, the factors that capture firm constraints such as access to finance and electricity do not appear to explain the gender productivity gap. Our findings have clear implications for policy. Given the role played by firm characteristics in explaining the gender productivity gap, there is a need for training programmes for female entrepreneurs in maintaining accounts as well as in facilitating the registration of their enterprises with state authorities.

**References**


Acar, A.C. (1993). 'The Impact of Key Internal Factors on Firm Performance: An Empirical Study of Small Turkish Firms'. *Journal of Small Business Management*, 31(4): 86–92.

Ahmed, S., and P. Maitra (2015). 'A Distributional Analysis of the Gender Wage Gap in Bangladesh'. *Journal of Development Studies*, 51(11): 1444–58. https://doi.org/10.1080/00220388.2015.1046444

Aterido, R., M. Hallward-Driemeier, and C. Pages (2011). 'Big Constraints to Small Firms' Growth? Business Environment and Employment Growth across Firms'. *Economic Development and Cultural Change*, 59(3): 609–47. https://doi.org/10.1086/658349

Autio, E. (2005). 'Toward a Theory of International New Ventures'. *Journal of International Business Studies*, 36(1): 9–19. https://doi.org/10.1057/palgrave.jibs.8400117

Ayyagari, Meghana, A. Demirguc-Kunt, and V. Maksimovic (2014). 'Who Creates Jobs in Developing Countries?'. *Small Business Economics*, 43(1): 75–99. https://doi.org/10.1007/s11187-014-9549-5

Banerji, S., R.S.N. Raj, and K. Sen (2016). 'Monitoring Costs, Credit Constraints and Entrepreneurship'. *The Manchester School*, 84(5): 573–99. https://doi.org/10.1111/manc.12122

Bardasi, E., S. Sabarwal, and K. Terrell (2011). 'How Do Female Entrepreneurs Perform? Evidence from Three Developing Regions'. *Small Business Economics*, 37(4): 417–41. https://doi.org/10.1007/s11187-011-9374-z

Barsky, R., J. Bound, K. Charles, and J. Lupton (2002). 'Accounting for the Black–White Wealth Gap: A Nonparametric Approach'. *Journal of the American Statistical Association*, 97: 663–73. https://doi.org/10.1198/016214502388618401

Batista, C., S. Sequeira, and P. Vicente (2021). 'Closing the Gender Profit Gap'. Mimeo. https://doi.org/10.2139/ssrn.3803152

Berry, A., E. Rodrigues, and H. Sandee (2002). 'Firm and Group Dynamics in the Small and Medium Enterprise Sector in Indonesia'. *Small Business Economics*, 18(1–3): 141–61. https://doi.org/10.1023/A:1015186023309

Bertrand, M. (2020). 'Gender in the 21st Century'. *AEA Papers and Proceedings*, 110: 1–24, Richard T. Ely Lecture. https://doi.org/10.1257/pandp.20201126

Bigsten, A., and M. Gebreeyesus (2007). 'The Small, the Young, and the Productive: Determinants of Manufacturing Firm Growth in Ethiopia'. *Economic Development and Cultural Change*, 55(4): 813–40. https://doi.org/10.1086/516767

Binks M.R., and C.T. Ennew (1996). 'Growing Firms and the Credit Constraint'. *Small Business Economics*, 8(1): 17–25. https://doi.org/10.1007/BF00391972

Bollard, A., P.J. Klenow, and H. Li (2014). 'Entry Costs Rise with Development'. Unpublished manuscript, Stanford University.

Borrowman, M., and S. Klasen (2020). 'Drivers of Gendered Sectoral and Occupational Segregation in Developing Countries'. *Feminist Economics*, 26(2): 62–94. https://doi.org/10.1080/13545701.2019.1649708

Calvo, J.L. (2006). 'Testing Gibrat's Law for Small, Young and Innovating Firms'. *Small Business Economics*, 26: 117–23. https://doi.org/10.1007/s11187-004-2135-5

Campos, F., and M. Gassier (2015). 'What Are the Constraints to Women-Owned Firms, What Works, What Doesn't, and What We Don't Know'. Policy Research Working Paper. Washington, DC: World Bank.

Carranza, E., C. Dhakal, and I. Love (2018). 'Female Entrepreneurs: How and Why Are They Different?'. Jobs Working Paper 20. Washington, DC: World Bank. https://doi.org/10.1596/31004

Chatterjee, S., T. Lebesmuehlbacher, and A. Narayanan (2021). 'How Productive Is Public Investment? Evidence from Formal and Informal Production in India'. *Journal of Development Economics*, 151: Article 102625. https://doi.org/10.1016/j.jdeveco.2021.102625

Chaudhuri, K., S. Sasidharan, and R.S.N. Raj (2020). 'Gender, Small Firm Ownership, and Credit Access: Some Insights from India'. *Small Business Economics*, 54: 1165–81. https://doi.org/10.1007/s11187-018-0124-3

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013). 'Inference on Counterfactual Distributions'. *Econometrica*, 81: 2205–68. https://doi.org/10.3982/ECTA10582

Coad, A., A. Segarra-Blascoand, and M. Teruel (2013). 'Like Milk or Wine: Does Firm Performance Improve with Age?'. *Structural Change and Economic Dynamics*, 24: 173–89. https://doi.org/10.1016/j.strueco.2012.07.002

Cotton, J. (1988). 'On the Decomposition of Wage Differentials'. *Review of Economics and Statistics*, 70: 236–43. https://doi.org/10.2307/1928307

Davino, C., M. Furno, and D. Vistocco (2013). *Quantile Regression: Theory and Applications*. New York: Wiley. https://doi.org/10.1002/9781118752685

De Mel, S., D. McKenzie, and C. Woodruff (2009). 'Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns'. *American Economic Journal: Applied Economics*, 1(3): 1–32. https://doi.org/10.1257/app.1.3.1

Delecourt, S., and O. Ng (2021). 'Does Gender Matter for Small Business Performance? Experimental Evidence from India'. Mimeo, Stanford University. https://doi.org/10.2139/ssrn.3830935

Deshpande, A., and S. Sharma (2013). 'Entrepreneurship or Survival? Caste and Gender of Small Business in India'. *Economic and Political Weekly*, 28(28): 38–49.

Deshpande, A., and S. Sharma (2016). 'Disadvantage and Discrimination in Self-Employment: Caste Gaps in Earnings in Indian Small Businesses'. *Small Business Economics*, 46: 325–46. https://doi.org/10.1007/s11187-015-9687-4

Dunne, T., M.J. Roberts, and L. Samuelson (1989). 'The Growth and Failure of U.S. Manufacturing Plants'. *Quarterly Journal of Economics*, 104: 671–98. https://doi.org/10.2307/2937862

Ericson, R., and A. Pakes (1995). 'Markov-Perfect Industry Dynamics: A Framework for Empirical Work'. *Review of Economic Studies*, 62: 53–82. https://doi.org/10.2307/2297841

Fariñas, J.C., and L. Moreno (2000). 'Firms' Growth, Size and Age: A Nonparametric Approach'. *Review of Industrial Organization*, 17(3): 249–65. https://doi.org/10.1023/A:1007834210622

Ferreira, F.H.G., S.P. Firpo, and J. Messina (2021). 'Labor Market Experience and Falling Earnings Inequality in Brazil: 1995–2012'. *The World Bank Economic Review*, article lhab005. https://doi.org/10.1093/wber/lhab005

Firpo, S., N.M. Fortin, and T. Lemieux (2007). 'Decomposing Wage Distributions Using Recentered Influence Functions Regressions'. Mimeo. University of British Columbia, Vancouver.

Firpo, S., N.M. Fortin, and T. Lemieux (2009). 'Unconditional Quantile Regressions'. *Econometrica*, 77: 953–73. https://doi.org/10.3982/ECTA6822

Firpo, S., N.M. Fortin, and T. Lemieux (2018). 'Decomposing Wage Distributions Using Recentered Influence Function Regressions'. *Econometrics*, 6(2): 28. https://doi.org/10.3390/econometrics6020028

Fortin, N., T. Lemieux, and S. Firpo (2011). 'Decomposition Methods in Economics'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, vol. IV.A, pp. 1–102. Amsterdam: North-Holland. https://doi.org/10.1016/S0169-7218(11)00407-2

Gang, I.N., R.S.N. Raj, and K. Sen (2020). 'Finance, Gender, and Entrepreneurship India's Informal Sector Firms'. WIDER Working Paper 2020/144. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/901-3

Gardeazabal, J., and A. Ugidos (2004). 'More on the Identification in Detailed Wage Decompositions'. *Review of Economics and Statistics*, 86: 1034–57. https://doi.org/10.1162/0034653043125239

Ghani, E., W.R. Kerr, and S.D. O'Connell (2014a). 'Political Reservations and Women's Entrepreneurship in India'. *Journal of Development Economics*, 108: 138–53. https://doi.org/10.1016/j.jdeveco.2014.01.008

Ghani, E., W.R. Kerr, and S.D. O'Connell (2014b). 'Spatial Determinants of Entrepreneurship in India'. *Regional Studies*, 48(6): 1071–89. https://doi.org/10.1080/00343404.2013.839869

Ginther, D.K., and K.J. Hayes (2003). 'Gender Differences in Salary and Promotion for Faculty in the Humanities, 1977–1995'. *The Journal of Human Resources*, 38(1): 34–73. https://doi.org/10.3368/jhr.XXXVIII.1.34

Goedhuys, M., and L. Sleuwaegen (2000). 'Entrepreneurship and Growth of Entrepreneurial Firms in Côte d'Ivoire'. *The Journal of Development Studies*, 36(3): 123–45. https://doi.org/10.1080/00220380008422631

Goraya, S.S. (2019). 'How Does Caste Affect Entrepreneurship? Birth vs Worth'. Mimeo, Universitat Pompeu Fabra.

Grimm, M., P. Knorringa, and J. Lay (2012). 'Constrained Gazelles: High Potentials in West Africa's Informal Economy'. *World Development*, 40(7): 1352–68. https://doi.org/10.1016/j.worlddev.2012.03.009

Guérin, I., S. Michiels, and G. Venkatasubramanian (2015). 'Labour in Contemporary South India'. In B. Harriss-White and J. Heyer (eds), *Indian Capitalism in Development*, pp. 118–36. London: Routledge. https://doi.org/10.4324/9781315770963-7

Hadlock, C.J., and J.R. Pierce (2010). 'New Evidence on Measuring Financial Constraints: Moving beyond the KZ Index'. *Review of Financial Studies*, 23: 1909–40. https://doi.org/10.1093/rfs/hhq009

Harriss-White, B., K. Vidyarthee, and A. Dixit (2014). 'Regions of Dalit and Adivasi Discrimination in India's Business Economy'. In B. Harriss-White and E. Basile (eds), *Dalits and Adivasis in India's Business Economy: Three Essays and an Atlas*. Gurgaon, India: Three Essays Collective. https://doi.org/10.1177/2394481120140106

Hsieh, C., and P.J. Klenow (2014). 'The Life Cycle of Plants in India and Mexico'. *The Quarterly Journal of Economics*, 129(3): 1035–84. https://doi.org/10.1093/qje/qju014

Jann, B. (2008). 'The Blinder–Oaxaca Decomposition for Linear Regression Models'. *The Stata Journal*, 8(4): 453–79. https://doi.org/10.1177/1536867X0800800401

Kathuria, V., R.S.N. Raj, and K. Sen (2013). 'The Effects of Economic Reforms on Manufacturing Dualism: Evidence from India'. *Journal of Comparative Economics*, 41(4): 1240–62. https://doi.org/10.1016/j.jce.2012.10.003

Kesar, S., and S. Bhattacharya (2019). 'Dualism and Structural Transformation: The Informal Manufacturing Sector in India'. *European Journal of Development Research*, 50: 560–82. https://doi.org/10.1057/s41287-019-00228-0

Khurana, S., and K. Mahajan (2020). 'Evolution of Wage Inequality in India (1983–2017): The Role of Occupational Task Content'. WIDER Working Paper 2020/167. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/924-2

Liedholm, C., and D. Mead (1999). *Small Enterprises and Economic Development: The Dynamic Role of Micro and Small Enterprises*. London: Routledge.

Marlow, S., and M. McAdam (2013). 'Gender and Entrepreneurship: Advancing Debate and Challenging Myths; Exploring the Mystery of the Under-Performing Female Entrepreneur'. *International Journal of Entrepreneurial Behavior & Research*, 19(1): 114–24. https://doi.org/10.1108/13552551311299288

Mazumdar, D., and S. Sarkar (2013). *Manufacturing Enterprises in Asia: Size Structure and Economic Growth*. London and New York: Routledge.

Mosse, D. (2018). 'Caste and Development: Contemporary Perspectives on a Structure of Discrimination and Advantage'. *World Development*, 110: 422–36. https://doi.org/10.1016/j.worlddev.2018.06.003

Nataraj, S. (2011). 'The Impact of Trade Liberalization on Productivity: Evidence from India's Formal and Informal Manufacturing Sectors'. *Journal of International Economics*, 85(2): 292–301. https://doi.org/10.1016/j.jinteco.2011.07.003

Neumark, D. (1988). 'Employers' Discriminatory Behavior and the Estimation of Wage Discrimination'. *Journal of Human Resources*, 23: 279–95. https://doi.org/10.2307/145830

NSSO (2017). 'Key Indicators of Unincorporated Non-agricultural Enterprises (Excluding Construction) in India'. NSS 73rd Round, July 2015–June 2016, Ministry of Statistics and Programme Implementation, Government of India.

Oaxaca, R. (1973). 'Male–Female Wage Differentials in Urban Labor Markets'. *International Economic Review*, 14: 693–709. https://doi.org/10.2307/2525981

Oaxaca, R.L. (2007). 'The Challenge of Measuring Labor Market Discrimination against Women'. *Swedish Economic Policy Review*, 14(1): 199–231.

Ojah, K., T. Gwatidzo, and S. Kaniki (2010). 'Legal Environment, Finance Channels and Investment: The East African Example'. *Journal of Development Studies*, 46: 724–44. https://doi.org/10.1080/00220380903012722

Oliveira, B., and A. Fortunato (2006). 'Firm Growth and Liquidity Constraints: A Dynamic Analysis'. *Small Business Economics*, 27(2): 139–56. https://doi.org/10.1007/s11187-006-0006-y

Penrose, E.G. (1959). *The Theory of the Growth of the Firm*. New York: Wiley.

Poschke, M. (2018). 'The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology'. *American Economic Journal: Macroeconomics*, 10: 1–41. https://doi.org/10.1257/mac.20140181

Raj, R.S.N., and S. Sasidharan (2014). 'The Growth Barriers of Informal Sector Enterprises: Evidence from India'. *The Developing Economies*, 52(4): 351–75. https://doi.org/10.1111/deve.12057

Raj, R.S.N., and K. Sen (2012). 'Did International Trade Destroy or Create Jobs in Indian Manufacturing?'. *European Journal of Development Research*, 24(3). https://doi.org/10.1057/ejdr.2011.19

Raj, R.S.N., and K. Sen (2015). 'Finance Constraints and Firm Transition in the Informal Sector: Evidence from Indian Manufacturing'. *Oxford Development Studies* 43(1): 123–43. https://doi.org/10.1080/13600818.2014.972352

Raj, R.S.N., and K. Sen (2016a). *Out of the Shadows? The Informal Manufacturing in Post-Reform India*. Delhi: Oxford University Press.

Raj, R.S.N., and K. Sen (2016b). 'Some Puzzles about Firms: Analyses of India's Informal Manufacturing Sector'. *Economic and Political Weekly*, 51(7): 43–51.

Rajan, R., and L. Zingales (1998). 'Financial Dependence and Growth'. *American Economic Review*, 88(3): 559–86.

Reimers, C.W. (1983). 'Labor Market Discrimination against Hispanic and Black Men'. *Review of Economics and Statistics*, 65: 570–79. https://doi.org/10.2307/1935925

Rios-Avila, F. (2020). 'Recentered Influence Functions (RIFs) in Stata: RIF Regression and RIF Decomposition'. *The Stata Journal*, 20(1): 51–94. https://doi.org/10.1177/1536867X20909690

Sharma, S (2014). 'Benefits of a Registration Policy for Microenterprise Performance in India'. *Small Business Economics*, 42(1): 153–64. https://doi.org/10.1007/s11187-013-9475-y

Sher, P.J., and P.Y. Yang (2005). 'The Effects of Innovative Capabilities and R&D Clustering on Firm Performance: The Evidence of Taiwan's Semiconductor Industry'. *Technovation*, 25(1): 1–82. https://doi.org/10.1016/S0166-4972(03)00068-3

Sleuwaegen, L., and M. Goedhuys (2002). 'Growth of Firms in Developing Countries: Evidence from Côte d'Ivoire'. *Journal of Development Economics*, 68(1): 117–35. https://doi.org/10.1016/S0304-3878(02)00008-1

Thorat, S., and N. Sadana (2009). 'Caste and Ownership of Private Enterprises'. *Economic and Political Weekly*, 44(23): 13–16.

Winker, P. (1999). 'Causes and Effects of Financing Constraints at the Firm Level'. *Small Business Economics*, 12(2): 169–81. https://doi.org/10.1023/A:1008035826914

World Bank (2012). *World Development Report 2012: Gender Equality and Development*. Washington, DC: World Bank. https://doi.org/10.1596/978-0-8213-9575-2

Yang C.H., and C.H. Huang (2005). 'R&D, Size and Firm Growth in Taiwan's Electronics Industry'. *Small Business Economics*, 25(5): 477–87. https://doi.org/10.1007/s11187-004-6487-7

Yun, M.-S. (2005). 'A Simple Solution to the Identification Problem in Detailed Wage Decompositions'. *Economic Inquiry*, 43: 766–72. https://doi.org/10.1093/ei/cbi053

## Appendix

Table A1: Decomposing productivity gap by percentiles: RIF-Oaxaca detailed decomposition

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
| Overall | | | | | | | | | |
| Mean male labour productivity (M) | 8.9267*** | 9.4879*** | 9.8341*** | 10.0834*** | 10.2867*** | 10.4887*** | 10.6893*** | 10.9173*** | 11.2318*** |
| | (0.0462) | (0.0356) | (0.0307) | (0.0268) | (0.0262) | (0.0264) | (0.0257) | (0.0282) | (0.0279) |
| Mean female labour productivity (F) | 7.8367*** | 8.3412*** | 8.6412*** | 8.8968*** | 9.1454*** | 9.4095*** | 9.6778*** | 10.0169*** | 10.4740*** |
| | (0.0976) | (0.0751) | (0.0829) | (0.0639) | (0.0949) | (0.0888) | (0.0788) | (0.0984) | (0.1217) |
| Gender gap in productivity (M−F) | 1.0900*** | 1.1467*** | 1.1929*** | 1.1866*** | 1.1414*** | 1.0793*** | 1.0115*** | 0.9003*** | 0.7578*** |
| | (0.0842) | (0.0655) | (0.0755) | (0.0576) | (0.0875) | (0.0797) | (0.0694) | (0.0881) | (0.1112) |
| Reweighting decomposition | | | | | | | | | |
| Counterfactual (C) | 8.6254*** | 9.1941*** | 9.5043*** | 9.7797*** | 10.0183*** | 10.2414*** | 10.4537*** | 10.7006*** | 11.0003*** |
| | (0.0558) | (0.0378) | (0.0408) | (0.0413) | (0.0374) | (0.0322) | (0.0341) | (0.0307) | (0.0299) |
| Total composition effect (M−C) | 0.3013*** | 0.2938*** | 0.3299*** | 0.3037*** | 0.2685*** | 0.2473*** | 0.2356*** | 0.2166*** | 0.2315*** |
| | (0.0530) | (0.0382) | (0.0384) | (0.0351) | (0.0301) | (0.0264) | (0.0277) | (0.0262) | (0.0258) |
| Total structural effect (C−F) | 0.7887*** | 0.8529*** | 0.8630*** | 0.8829*** | 0.8729*** | 0.8319*** | 0.7759*** | 0.6837*** | 0.5263** |
| | (0.2202) | (0.1658) | (0.1735) | (0.1146) | (0.1862) | (0.1635) | (0.1469) | (0.1970) | (0.2580) |
| RIF aggregate decomposition | | | | | | | | | |
| Pure composition effect | 0.3702*** | 0.3554*** | 0.3103*** | 0.2636*** | 0.2325*** | 0.2133*** | 0.1905*** | 0.1995*** | 0.2059*** |
| | (0.1375) | (0.0710) | (0.0490) | (0.0339) | (0.0321) | (0.0262) | (0.0236) | (0.0218) | (0.0273) |
| Specification error | −0.0689 | −0.0616 | 0.0196 | 0.0401 | 0.0359 | 0.0340 | 0.0451* | 0.0171 | 0.0256 |
| | (0.1525) | (0.0800) | (0.0557) | (0.0364) | (0.0331) | (0.0281) | (0.0242) | (0.0233) | (0.0257) |
| Pure structural effect | 0.7921*** | 0.8499*** | 0.8461*** | 0.8675*** | 0.8559*** | 0.8096*** | 0.7483*** | 0.6585*** | 0.5002* |
| | (0.2213) | (0.1670) | (0.1720) | (0.1117) | (0.1821) | (0.1597) | (0.1427) | (0.1931) | (0.2553) |
| Reweighting error | −0.0034 | 0.0030 | 0.0169 | 0.0154 | 0.0170 | 0.0223 | 0.0275 | 0.0251 | 0.0261 |
| | (0.0341) | (0.0262) | (0.0327) | (0.0322) | (0.0298) | (0.0261) | (0.0267) | (0.0226) | (0.0211) |
| Pure composition effect | | | | | | | | | |
| Firm characteristics | 0.2044** | 0.1918*** | 0.1677*** | 0.1369*** | 0.1242*** | 0.1114*** | 0.1033*** | 0.1000*** | 0.0901*** |
| | (0.0830) | (0.0441) | (0.0311) | (0.0227) | (0.0205) | (0.0184) | (0.0164) | (0.0143) | (0.0196) |
| Firm constraints | 0.0030 | 0.0007 | 0.0001 | −0.0008 | −0.0011 | −0.0008 | −0.0007 | −0.0004 | −0.0011 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (0.0049) | (0.0036) | (0.0030) | (0.0023) | (0.0020) | (0.0018) | (0.0017) | (0.0017) | (0.0017) |
| Social group | 0.0279*** | 0.0178*** | 0.0131** | 0.0100** | 0.0078* | 0.0077* | 0.0067* | 0.0061 | 0.0054 |
| | (0.0106) | (0.0066) | (0.0053) | (0.0047) | (0.0042) | (0.0046) | (0.0037) | (0.0043) | (0.0044) |
| Regional effects | 0.0203 | 0.0210 | 0.0155 | 0.0186 | 0.0165 | 0.0204 | 0.0192 | 0.0233 | 0.0322** |
| | (0.0504) | (0.0343) | (0.0257) | (0.0205) | (0.0177) | (0.0168) | (0.0151) | (0.0160) | (0.0139) |
| Sectoral effects | 0.1146** | 0.1240*** | 0.1140*** | 0.0989*** | 0.0851*** | 0.0746*** | 0.0621*** | 0.0706*** | 0.0794*** |
| | (0.0524) | (0.0319) | (0.0217) | (0.0160) | (0.0151) | (0.0129) | (0.0134) | (0.0120) | (0.0165) |
| Pure structural effect | | | | | | | | | |
| Firm characteristics | 0.1104 | −0.2090 | −0.1936 | −0.2134 | −0.0115 | 0.0545 | 0.0923 | 0.1640 | 0.5261 |
| | (0.9169) | (0.5380) | (0.5074) | (0.3684) | (0.4608) | (0.4615) | (0.4500) | (0.4380) | (0.6057) |
| Firm constraints | 0.1577 | −0.1072 | −0.0446 | 0.0087 | 0.0588 | 0.0777 | 0.0741 | 0.0828 | 0.0648 |
| | (0.4716) | (0.5127) | (0.4227) | (0.2037) | (0.2364) | (0.2782) | (0.2132) | (0.2538) | (0.2829) |
| Social group | 0.2008 | 0.1040 | 0.0469 | 0.0767 | 0.0287 | 0.0272 | 0.0225 | 0.0284 | −0.0340 |
| | (0.1745) | (0.1364) | (0.1323) | (0.1078) | (0.1509) | (0.1598) | (0.1275) | (0.1578) | (0.0683) |
| Regional effects | 0.0111 | −0.0282 | 0.0532 | 0.0365 | 0.1713 | 0.2092 | 0.1176 | 0.2523 | 0.3367 |
| | (0.2437) | (0.1753) | (0.1703) | (0.1313) | (0.1716) | (0.1935) | (0.1523) | (0.1948) | (0.2139) |
| Sectoral effects | 0.0629 | 0.2796 | 0.2501 | 0.1739 | 0.1994 | 0.2312 | 0.2640 | 0.0638 | 0.0611 |
| | (0.3482) | (0.2463) | (0.2212) | (0.1729) | (0.2221) | (0.2393) | (0.2732) | (0.4263) | (0.6629) |
| Intercept | 0.2491 | 0.8107 | 0.7342 | 0.7851* | 0.4093 | 0.2097 | 0.1778 | 0.0672 | −0.4545 |
| | (1.3414) | (0.6533) | (0.5679) | (0.4559) | (0.5949) | (0.5983) | (0.5831) | (0.6803) | (0.9268) |
| No. of observations | 270,442 | 270,442 | 270,442 | 270,442 | 270,442 | 270,442 | 270,442 | 270,442 | 270,442 |

Note: the dependent variable is the estimated RIF at the respective decile of the log labour productivity. The gender gap is the difference between the log labour productivity of male-owned and female-owned enterprises. Sampling weights are used in estimations. Standard errors reported in parentheses are robust to heteroskedasticity and clustered residuals within districts. The reweighting factors are estimated using a logit model. ***, **, and * denote significance at the 1, 5, and 10 per cent levels, respectively. See Table 5 for details of the variables included in each category.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).

Table A2: Decomposition results: Percentile ratios, Gini coefficient, and variance

| Variables | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | iqr9010 | iqr5010 | iqr9050 | Gini | Var |
| Overall | | | | | |
| Mean male labour productivity (M) | 2.3051*** | 1.3600*** | 0.9451*** | 0.4402*** | 0.9382*** |
| | (0.0414) | (0.0333) | (0.0149) | (0.0051) | (0.0335) |
| Mean female labour productivity (F) | 2.6370*** | 1.3086*** | 1.3284*** | 0.5161*** | 1.1223*** |
| | (0.1368) | (0.1159) | (0.1024) | (0.0121) | (0.0800) |
| Gender gap in productivity (M−F) | −0.3319*** | 0.0514 | −0.3833*** | −0.0759*** | −0.1841** |
| | (0.1287) | (0.1119) | (0.1010) | (0.0110) | (0.0763) |
| Reweighting decomposition | | | | | |
| Counterfactual (C) | 2.3749*** | 1.3929*** | 0.9820*** | 0.4411*** | 0.9555*** |
| | (0.0434) | (0.0375) | (0.0168) | (0.0046) | (0.0271) |
| Total composition effect (M−C) | −0.0698 | −0.0329 | −0.0370* | −0.0009 | −0.0172 |
| | (0.0510) | (0.0428) | (0.0202) | (0.0069) | (0.0439) |
| Total structural effect (C−F) | −0.2621 | 0.0842 | −0.3464 | −0.0750*** | −0.1668 |
| | (0.3102) | (0.2680) | (0.2462) | (0.0263) | (0.1786) |
| RIF aggregate decomposition | | | | | |
| Pure composition effect | −0.1643 | −0.1376 | −0.0266 | −0.0125 | −0.0937 |
| | (0.1423) | (0.1336) | (0.0414) | (0.0124) | (0.1145) |
| Specification error | 0.0944 | 0.1048 | −0.0103 | 0.0116 | 0.0765 |
| | (0.1582) | (0.1465) | (0.0413) | (0.0137) | (0.1237) |
| Pure structural effect | −0.2917 | 0.0638 | −0.3555 | −0.0808*** | −0.2082 |
| | (0.3112) | (0.2680) | (0.2461) | (0.0265) | (0.1772) |
| Reweighting error | 0.0295 | 0.0204 | 0.0091 | 0.0058* | 0.0414** |
| | (0.0246) | (0.0204) | (0.0120) | (0.0034) | (0.0195) |
| Pure composition effect | | | | | |
| Firm characteristics | −0.1143 | −0.0802 | −0.0341 | −0.0100 | −0.0534 |
| | (0.0870) | (0.0810) | (0.0277) | (0.0089) | (0.0551) |
| Firm constraints | −0.0041 | −0.0040 | −0.0001 | −0.0005 | −0.0018 |
| | (0.0050) | (0.0048) | (0.0022) | (0.0006) | (0.0053) |
| Social group | −0.0225*** | −0.0201** | −0.0024 | −0.0018** | −0.0228* |
| | (0.0086) | (0.0081) | (0.0019) | (0.0007) | (0.0125) |

| | | | | | |
|---|---|---|---|---|---|
| Regional effects | 0.0119 | −0.0038 | 0.0157 | 0.0017 | 0.0131 |
| | (0.0495) | (0.0426) | (0.0130) | (0.0054) | (0.0376) |
| Sectoral effects | −0.0353 | −0.0295 | −0.0058 | −0.0019 | −0.0288 |
| | (0.0563) | (0.0499) | (0.0204) | (0.0058) | (0.0562) |
| Pure structural effect | | | | | |
| Firm characteristics | 0.3302 | −0.1390 | 0.4693 | 0.1543 | 0.3003 |
| | (1.1308) | (1.0604) | (0.7059) | (0.1560) | (0.5603) |
| Firm constraints | −0.0954 | −0.1013 | 0.0060 | 0.0118 | 0.2697 |
| | (0.6108) | (0.5443) | (0.3124) | (0.0680) | (0.5593) |
| Social group | −0.2257 | −0.1680 | −0.0577 | −0.0201 | −0.1374 |
| | (0.1874) | (0.2154) | (0.1520) | (0.0223) | (0.1074) |
| Regional effects | 0.2735 | 0.1486 | 0.1249 | 0.0643 | 0.2138 |
| | (0.2839) | (0.2715) | (0.2195) | (0.1057) | (0.1962) |
| Sectoral effects | −0.0144 | 0.1349 | −0.1494 | −0.0489 | 0.0118 |
| | (0.6226) | (0.2852) | (0.4709) | (0.0484) | (0.2981) |
| Intercept | −0.5599 | 0.1886 | −0.7485 | −0.2422 | −0.8664 |
| | (1.5955) | (1.4989) | (0.9031) | (0.1872) | (0.7520) |
| No. of observations | 270,442 | 270,442 | 270,442 | 270,442 | 270,442 |

Note: standard errors reported in parentheses are robust to heteroskedasticity and clustered residuals within districts. Sampling weights are used in estimations. The reweighting factors are estimated using a logit model. ***, **, and * denote significance at the 1, 5, and 10 per cent levels, respectively. The iqr9010 is the difference between the 90th percentile and the 10th percentile of log labour productivity. The series iqr5010 and iqr9050 are computed analogously. The Gini coefficient is expressed in percentage points and ranges from 0 (perfect equality) to 100 (perfect inequality). Variance captures the impact of covariates on the variance of the distributions of log labour productivity. See Table 5 for details of the variables included in each category.

Source: authors' calculations based on NSSO survey data on unincorporated non-agricultural enterprises, 73rd round (2015/16).